



UNIVERSIDAD PERUANA
CAYETANO HEREDIA

FACULTAD DE CIENCIAS E INGENIERÍA

**“PREDICCIÓN *IN SILICO* DE PROTEÍNAS DE SECRECIÓN DE
MYCOBACTERIUM TUBERCULOSIS H37RV DIRIGIDAS AL NÚCLEO DE
LAS CÉLULAS HUMANAS Y SU INTERACCIÓN CON EL ADN”**

AUTORES

Bach. Marco Orlando Acosta Rincon

Bach. Liani Sairith Bustamante Gonza

Tesis para optar el Título profesional de Licenciado en Biología

ASESORA

PhD.MSc. Claudia Machicado Rivero

Lima- Perú

2023

"PREDICCIÓN IN SILICO DE PROTEINAS DE SECRECIÓN DE MYCOBACTERIUM TUBERCULOSIS H37RV DIRIGIDAS AL NÚCLEO DE LAS CÉLULAS HUMANAS Y SU INTERACCIÓN CON EL ADN"

INFORME DE ORIGINALIDAD

6%

INDICE DE SIMILITUD

5%

FUENTES DE INTERNET

4%

PUBLICACIONES

0%

TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1

hdl.handle.net

Fuente de Internet

1%

2

Hernández Miranda Olga Andrea. "Análisis transcriptómico diferencial del metabolismo de auxinas en ovarios, durante la transición de flor a fruto en dos genotipos de Vanilla planifolia Andrews (Orchidaceae)", TESIUNAM, 2021

Publicación

<1%

3

idoc.pub

Fuente de Internet

<1%

4

www.colibri.udelar.edu.uy

Fuente de Internet

<1%

5

Méndez Sánchez Lucia. "Fraccionamiento de biomoléculas del complejo Mycobacterium tuberculosis", TESIUNAM, 2009

Publicación

<1%

6

Submitted to Universidad Peruana Cayetano Heredia

Trabajo del estudiante

<1%

Contenido

RESUMEN	1
ABSTRACT	2
1. INTRODUCCIÓN	3
1.1. Microorganismos causantes de cáncer	3
1.1.1. Bacterias y cáncer	3
1.1.2. Otros microorganismos vinculados con el cáncer	5
1.2. <i>Mycobacterium tuberculosis</i>	5
1.2.1. Infección por <i>M. tuberculosis</i>	5
1.2.2. Cáncer de pulmón (CP) asociado con tuberculosis previa	6
1.2.3. Daño en el ADN por infección por <i>M. tuberculosis</i>	7
1.2.4. Proteoma y secretoma de <i>M. tuberculosis</i>	7
1.3. Predicción de proteínas bacterianas de secreción, dirigidas al núcleo de las células huésped.....	8
2. PREGUNTA DE INVESTIGACIÓN Y OBJETIVOS	9
2.1. Descripción del problema	9
2.2. Justificación	11
2.3. Pregunta de investigación.....	12
2.4. Objetivo general	12
2.5. Objetivos específicos	12
3. MATERIALES Y MÉTODOS	13
3.1. Diseño del estudio.....	13
3.2. Obtención de la base de datos del proteoma y secretoma de <i>M. tuberculosis</i>	13
3.3. Determinación <i>in silico</i> de proteínas de secreción.....	14
3.4. Predicción de la localización subcelular en las células huésped.	14
3.5. Análisis de las propiedades fisicoquímicas de las proteínas de localización nuclear.....	15
3.6. Ontología génica, exploración de dominios proteicos y predicción de unión al ADN.....	15
3.7. Determinación de homología con genotoxinas bacterianas.....	16
3.8. Modelamiento y visualización molecular	16
3.9. Docking molecular	17
3.10. Cuestiones éticas.....	18
4. RESULTADOS	18

4.1.	Compilación del proteoma de <i>M. tuberculosis</i>	19
4.2.	Predicción de las proteínas de secreción	20
4.3.	Localización subcelular de las proteínas de secreción.	22
4.4.	Propiedades fisicoquímicas: Peso molecular (PM) y punto isoelectrico (PI)	24
4.5.	Ontología génica, dominios proteicos y predicción de unión al ADN	27
4.6.	Estudio de homología con la genotoxina CdtB	37
4.7.	Modelamiento y visualización molecular	37
4.8.	<i>Docking</i> Molecular	41
5.	DISCUSIÓN	45
6.	CONCLUSIONES	52
7.	LIMITACIONES Y RECOMENDACIONES	52
	REFERENCIAS BIBLIOGRÁFICAS	54
	ANEXOS	65

RESUMEN

La infección pulmonar por *Mycobacterium tuberculosis* se asocia con el cáncer de pulmón. Así, las personas con historial de tuberculosis pulmonar tienen más riesgo de desarrollar cáncer de pulmón que aquellas que no tuvieron tuberculosis. Sin embargo, existe poca información sobre la participación de *M. tuberculosis* en la etiología de la neoplasia pulmonar. Se ha descrito que *M. tuberculosis* daña el ADN de las células humanas, aunque no se sabe cómo. Un posible mecanismo sería a través de proteínas secretadas como parte de su virulencia, que se dirijan al núcleo. Por ello, en este estudio nos propusimos predecir las proteínas secretadas por *M. tuberculosis* que se dirigen al núcleo de la célula huésped, mediante una difusión pasiva, y su potencial interacción con el ADN. Utilizando programas bioinformáticos, descubrimos que de las 972 proteínas que componen el secretoma de *M. tuberculosis*, 58 proteínas tienen como destino el núcleo de las células de su hospedador. Así mismo, de ellas, encontramos que 3 proteínas (la Proteína ribosómica 30S S3, la proteína reguladora transcripcional Rv3488 y la Formamidopirimidina-ADN glicosilasa) poseen una alta probabilidad de interactuar directamente con el ADN. Nuestros resultados respaldarían la hipótesis que vincula la infección por *M. tuberculosis* y el cáncer de pulmón, pudiendo tener, este grupo de proteínas de secreción destinadas al núcleo, potencial para alterar las funciones del genoma al interactuar con los ácidos nucleicos.

Palabras clave: *Mycobacterium tuberculosis*, proteínas de secreción, cáncer de pulmón, docking, simulación teórica.

ABSTRACT

Lung infection with *Mycobacterium tuberculosis* is associated with lung cancer. Thus, individuals with a history of pulmonary tuberculosis are at higher risk of developing lung cancer. However, there is little information on the involvement of *M. tuberculosis* in the etiology of lung neoplasia. It has been described that *M. tuberculosis* damages the DNA of human cells, although it is not known how. One possible mechanism would be through secreted proteins that target the nucleus as part of its virulence. In the present study, we set out to predict the proteins secreted by *M. tuberculosis* that target the host cell nucleus via passive diffusion, and their potential interaction with DNA. Using bioinformatics procedures, we discovered that of the 972 proteins that make up the *M. tuberculosis* secretome, 58 proteins target the host cell nucleus. Of these, 3 proteins (30S ribosomal protein S3, transcriptional regulatory protein Rv3488 and formamidopyrimidine-DNA glycosylase) have a high probability of interacting directly with DNA. Our results support the link between *M. tuberculosis* infection and lung cancer, and this group of secretion proteins destined to the nucleus may have the potential to alter genome functions by interacting with nucleic acids.

Keywords: *Mycobacterium tuberculosis*, secretion proteins, lung cancer, docking, theoretical simulation.

1. INTRODUCCIÓN

1.1. Microorganismos causantes de cáncer

El cáncer puede originarse por rasgos hereditarios o bien tener un origen somático, siendo este último el tipo más frecuente a nivel mundial (1). Cuando la célula normal se convierte en una célula cancerosa, bien sea por razones genéticas o somáticas, sufre una serie de alteraciones, proceso que recibe el nombre de transformación celular (2). La transformación de tipo somática puede ser causada tanto por los agentes carcinogénicos físicos (radiación UV e ionizante), como por compuestos químicos (componente del tabaco, arsénico, amianto, entre otros) y algunos agentes biológicos (2-4). Actualmente, existen 11 agentes biológicos (incluyendo bacterias, virus y parásitos) considerados por la Organización Mundial de la Salud (OMS) como carcinógenos de clase I (4). Estos estarían causando alrededor del 20% de los tumores malignos reportados en humanos (4).

1.1.1. Bacterias y cáncer

Actualmente, la única bacteria reconocida como agente carcinogénico por la OMS es *H. pylori* (3,5). Se sabe que la infección crónica por este microorganismo conlleva al desarrollo del cáncer de estómago (3,5). Sin embargo, estudios epidemiológicos han asociado otras infecciones bacterianas con el cáncer (1,4,6). Por ejemplo, se han relacionado las infecciones pulmonares por *Mycoplasma pneumoniae*, *Chlamydia pneumoniae* y *Mycobacterium tuberculosis* con el cáncer de pulmón (CP), asimismo, las cepas patogénicas de *Escherichia coli*, *Salmonella enteritidis*, *Streptococcus gallolyticus subsp. gallolyticus* (Sgg), *S. bovis* y *Bacteroides fragilis* se han vinculado con el cáncer colorrectal (1,4,6). También las infecciones por *C. trachomatis*, *M. genitalium* y *Neisseria gonorrhoeae* están vinculadas con el cáncer del tracto genitourinario; mientras que la infección por *Salmonella typhi* ha sido relacionada con el cáncer de vesícula biliar, y la de *Fusobacterium nucleatum* y *E. coli* con el cáncer esofágico (4,6). Además, la infección por *S. typhi* se ha vinculado con el cáncer de la vesícula biliar, y las infecciones por las bacterias de la familia de *Clostridia* y *Ruminococcaceae* se cree que podrían estar implicadas en la progresión del cáncer de mama (6). En este sentido, aún no se ha establecido la contribución real de estos agentes biológicos, ya que los mecanismos implicados pueden ser múltiples y variados (4).

1.1.1.1. Mecanismos carcinogénicos promovidos por bacterias

Las bacterias pueden inducir la transformación oncogénica a través de mecanismos directos o indirectos (1,3).

Uno de los mecanismos directos involucra la producción de factores de virulencia que se dirigen al núcleo de la célula huésped y dañan el ADN, tal como sucede con las genotoxinas bacterianas (1,4,7). Un segundo mecanismo sería la manipulación e interferencia de vías intracelulares importantes, por ejemplo, las rutas de reparación de daño al ADN (1,3). Otro mecanismo directo que desencadena la carcinogénesis, involucraría la alteración o desregulación de las vías de señalización que regulan el crecimiento, la apoptosis, la proliferación y la migración celular (1,7).

Por otro lado, uno de los mecanismos indirectos que poseen las bacterias e inducen la transformación, se relaciona con su capacidad de liberar especies reactivas de oxígeno (ROS), las cuales dañan al ADN provocando mutaciones en el material genético (1,7). En segundo lugar, mediante un mecanismo de supervivencia, las bacterias manipulan e inhiben la respuesta del sistema inmune, lo cual puede interferir en el control de la formación y el crecimiento de tumores (1,3). Por último, las bacterias pueden desencadenar un estado inflamatorio crónico que cause daño celular irreparable, tal como ocurre con *H. pylori*, la cual altera la división epitelial de la mucosa gástrica, conllevando al cáncer (1,5,7).

1.1.1.2. Toxinas bacterianas dirigidas al núcleo

Se sabe que algunas bacterias producen proteínas tóxicas que se dirigen al núcleo de las células huésped, donde inducen daño en el material genético, promoviendo así mutaciones y el cáncer (7,8). Estos compuestos son llamados genotoxinas y hasta la fecha, solo se ha determinado y caracterizado un número reducido de ellas (8). Entre dichos factores están las endonucleasas de *N. gonorrhoeae*, la toxina exoenzima S de *Pseudomonas aeruginosa*, la toxina shiga originada por *Shigella dysenteriae* y la proteína uropatógena específica (Usp) producida por bacterias como la cepa uropatógena de *E. coli* y *Salmonella bongori* (1,7-9).

Además, de las ya mencionadas, existen dos genotoxinas que han sido muy estudiadas. La primera de ellas es la toxina de distensión citoletal (CDT), producida por bacterias gram negativas como *E. coli*, *S. dysenteriae*, *Campylobacter jejuni*, *Aggregatibacter actinomycetemcomitans*, entre otras (10). Las células expuestas a una dosis subletal de CDT, sufren acumulación de mutaciones, ello debido a la ruptura del ADN causado por la subunidad

CdtB de esta genotoxina (8, 10). La segunda es la colibactina, generada por las cepas de *E. coli* del grupo filogenético B2 y por otras *Enterobacteriaceae* que presente el gen pks o Cld (11). Experimentos han mostrado que la colibactina causa roturas de la doble hebra del ADN (8,11). Asimismo, aquellas bacterias que producen colibactina han sido encontradas en las biopsias de pacientes con tumores colorrectales (8).

1.1.2. Otros microorganismos vinculados con el cáncer

Existen siete virus carcinógenos de acuerdo a la OMS, los cuales son: el virus de Epstein-Barr (VEB) causante de cáncer gástrico y linfoma, el virus de la hepatitis B y C (VHB y VHC) que ocasionan cánceres hepatocelulares, el virus del papiloma humano (VPH) que propicia el cáncer de ano y carcinomas cervicales, el virus linfotrópico T humano-1 (HTLV-1) que provoca leucemias, el herpes virus del sarcoma de Kaposi que causa tumores que afectan la piel y el virus de la inmunodeficiencia humana-1 que aumenta el riesgo de desarrollar linfomas no Hodgkin, cáncer de cuello uterino, entre otros (4,12).

Asimismo, existen tres parásitos carcinógenos que son los helmintos *Opisthorchis viverrini* y *Clonorchis sinensis* causantes de colangiocarcinomas, y *Schistosoma haematobium* causante del cáncer de vejiga (4, 13,14).

1.2. *Mycobacterium tuberculosis*

1.2.1. Infección por *M. tuberculosis*

La infección no controlada por *M. tuberculosis* causa la tuberculosis pulmonar, una enfermedad infecciosa considerada una de las principales causas de mortalidad en el mundo, siendo así un importante problema de salud pública (15). Se estima que 1,7 millones de personas están infectadas por esta micobacteria y que 10 millones de personas desarrollaron tuberculosis en el 2017, una enfermedad que causa el fallecimiento anual de alrededor de 1,45 millones de personas (15,16).

El ingreso de *M. tuberculosis* al organismo humano ocurre a través de la vía respiratoria superior para luego llegar al tracto respiratorio inferior, donde se puede internalizar e infectar las células alveolares (16,17). Esto desencadena una respuesta inflamatoria local que atrae a células inmunológicas, como las células dendríticas y los macrófagos, que lo fagocitan y se concentran en el sitio de infección, formando los granulomas (16,17). Dentro de las células inmunológicas, *M. tuberculosis* se encuentra en los fagosomas, vesículas que la transportan al

lisosoma para su eliminación (17). Sin embargo, la bacteria a través de diversos mecanismos y estrategias puede interrumpir la trayectoria de los fagosomas y residir en estos o liberarse e instalarse en el citosol de la célula donde empieza a multiplicarse hasta liberarse fuera de la célula (17). La micobacteria puede mantenerse en estado latente, en la que se presume que está inactiva y no se replica; sin embargo, también logra continuar con su desarrollo y multiplicación causando lesiones más extensas (16,17).

1.2.2. Cáncer de pulmón (CP) asociado con tuberculosis previa

El CP es una de las principales causas de muertes relacionadas con el cáncer, provocando el fallecimiento de 1.8 millones de personas por año (18). Entre los principales subtipos del CP reportados se encuentran el carcinoma de células escamosas, el carcinoma de células grandes y los adenocarcinomas (19). Existen varios factores que inducen al carcinoma de pulmón, en los que se incluye el consumo de tabaco y la exposición al gas de radón (19). Además, se piensa que las enfermedades pulmonares previas como la tuberculosis también podrían contribuir a la carcinogénesis pulmonar (19, 20). De hecho, se ha reportado la ocurrencia simultánea o secuencial de CP en personas con tuberculosis previa, por lo que ha resultado importante indagar si hay relación entre ambas patologías (18-22). En este sentido, diversos estudios han evaluado aquella posible relación. Por ejemplo, en una investigación se analizó 9 estudios epidemiológicos y se determinó que si existe una relación entre el CP y la tuberculosis (18). Otro estudio, sugiere que aquellas personas con antecedente de tuberculosis son más propensas a desarrollar CP, siendo las zonas del pulmón dañadas por la tuberculosis los posibles lugares donde se desarrolle la carcinogénesis (20). Asimismo, un estudio prospectivo de 16 años que se realizó a una cohorte poblacional coreana comprendida por más de 1 millón de personas adultas, encontró que la tuberculosis, tanto en fumadores como en no fumadores, se asocia con un mayor riesgo de incidencia de CP [Relación de riesgo (HR)=1,37 en hombres; HR=1. 49 en mujeres] (21). De igual modo, un estudio sistemático y de meta-análisis, determinó una asociación significativamente alta [Odds ratio (OR)= 2,09] entre la tuberculosis y el CP recién desarrollado (22). En conclusión, estos estudios sugieren que existe un riesgo significativamente mayor de desarrollar CP en personas con tuberculosis (18-22).

A pesar de contar con estudios de asociación, hasta la fecha, no se ha demostrado que la infección por *M. tuberculosis* sea una causa de CP ni tampoco los mecanismos moleculares que subyacen en esta relación infección-cáncer. Sin embargo, es posible que parte de dicho conocimiento pueda obtenerse de los estudios de proteómica.

1.2.3. Daño en el ADN por infección por *M. tuberculosis*

Motivados por los numerosos reportes sobre pacientes que tuvieron tuberculosis pulmonar y subsecuentemente CP (18,21,22), se han conducido investigaciones para determinar el poder carcinogénico de la micobacteria. Dichos estudios evaluaron si *M. tuberculosis* podía promover acumulación de mutaciones en el ADN en las células del huésped, condición conocida como inestabilidad genómica (23,24). La inestabilidad genómica antecede al cáncer, por lo que, si la infección por *M. tuberculosis* provoca inestabilidad del genoma, sería un fuerte indicio de su vínculo con la neoplasia.

Los resultados de dichas investigaciones indicaron que este patógeno inflige daño al ADN a través de mecanismos aún no dilucidados (23,24). El estudio de Castro-Garza et al. (23) fue un ensayo *in vitro* que usó la técnica de hibridación de fluorescencia *in situ* para detectar rotura del ADN (DBD-FISH). Se reportó que los macrófagos infectados por *M. tuberculosis* presentan inestabilidad genómica, como resultado de las roturas del ADN. Por su parte, en el estudio de Lochab S. et al. (24), se determinó que los macrófagos infectados por bacilos vivos de *M. tuberculosis* sufren un daño considerable de su ADN, lo cual fue determinado considerando los niveles γ H2AX, una proteína histona que se forma cuando hay daños en el material genético. Además, se sugiere que las proteínas secretadas por el sistema secretor accesorio SecA1 serían potencialmente las implicadas en el daño del ADN (24).

1.2.4. Proteoma y secretoma de *M. tuberculosis*

El proteoma se define como el conjunto de proteínas expresado por una determinada célula, que varía con la etapa de desarrollo y con el entorno en el que se encuentra (25). El proteoma de los microorganismos clínicamente relevantes ha sido estudiado con el fin de entender qué proteínas están involucradas en las afecciones que causan (25,26). Así, el proteoma de *M. tuberculosis*, incluyendo la cepa *H37Rv*, ha sido descrito, aplicando principalmente la espectrometría de masas (25-31). Un estudio de proteómica y glicoproteómica de la cepa *M. tuberculosis H37Rv* determinó la existencia de 1314 proteínas, mientras que otro, donde se evaluó la proteogenómica de la micobacteria, identificó alrededor de 3176 polipéptidos. La diferencia entre estos hallazgos se debe principalmente al uso de diferentes condiciones y estrategias de cultivo de la micobacteria y de técnicas usadas en la obtención de las proteínas (27,28). Así mismo, otro estudio ha determinado que 1051 proteínas están involucradas en la interacción huésped-patógeno, lo cual es importante porque tienen un rol fundamental en la

patogenicidad y sirven como marcadores diagnósticos, dianas terapéuticas y antígenos de vacunas (29).

También se han comparado los proteomas de diferentes cepas de *M. tuberculosis* (la cepa *H37Rv*, *W-Beijing*, *CAS* y *LAM3*), encontrando que 1938 polipéptidos son comunes a las cuatro cepas (30). Por otro lado, una característica de la micobacteria es su capacidad de permanecer en estado de latencia (donde la micobacteria está en un estado no replicante, ni transmisible) y luego reactivarse cuando las condiciones son óptimas (31). Estos estados biológicos también han sido caracterizados proteómicamente, determinándose que un total de 1871 proteínas son producidas tanto en el estado de latencia como el de reactivación (31). Por último, también se ha comparado el perfil proteómico de este patógeno cuando pasa de una fase logarítmica a un estado de latencia. Por lo que se sometieron a condiciones de presencia y de privación de nutrientes, imitando así la fase logarítmica y de latencia, respectivamente. De este modo, se identificó en ambas condiciones, un total de 1176 polipéptidos (32).

Por otro lado, el secretoma está constituido por un grupo de proteínas del proteoma que es liberado al espacio extracelular. Estas proteínas son relevantes en bacterias como *M. tuberculosis*, ya que intervienen en su virulencia y patogenia. El secretoma de *M. tuberculosis H37Rv*, también ha sido estudiado y determinado experimentalmente. Para ello se analizaron aislados clínicos, hallándose aproximadamente 837 proteínas secretoras de la micobacteria (33). Se sabe que estos polipéptidos tienen diversas funciones como la adhesión y regulación biológica, biogénesis de componentes celulares, metabolismo, respuesta inmunitaria, entre otras. Asimismo, se han aplicado técnicas bioinformáticas en la determinación del secretoma. Por ejemplo, en un estudio usando el programa SignalP 5.0 se predijo las proteínas de secreción de *M. tuberculosis H37Rv* a partir de su proteoma, identificándose 140 proteínas de secreción (27).

1.3. Predicción de proteínas bacterianas de secreción, dirigidas al núcleo de las células huésped

Como se ha mencionado anteriormente, es posible la participación de ciertas bacterias en la etiología del cáncer, por lo que diversas investigaciones buscan descubrir los mecanismos que se estarían suscitando. Por ejemplo, se cree que algunas bacterias como *C. pneumoniae*, *E. coli*, *S. typhimurium*, *Mycoplasma hominis* y *H. pylori* producen ciertos polipéptidos que podrían migrar a los diferentes orgánulos de la célula huésped, como el núcleo. La localización

nuclear de las proteínas bacterianas puede generar daño en el genoma y subsecuentemente alterar, e interferir en ciertas vías intracelulares, pudiendo propiciar incluso el cáncer (34-39).

Mediante estudios *in silico*, utilizando herramientas bioinformáticas, se han predicho proteínas de secreción en bacterias y su destino en la célula huésped, poniendo énfasis en el direccionamiento nuclear (34-39). Para predecir la localización subcelular en las células del huésped se han empleado programas que analizan las propiedades fisicoquímicas y la presencia de péptido señal, como mapper cNLS, BaCeILo, Hum-mPLoc 2.0 y LocTree3, (34-39). Los resultados obtenidos en las distintas investigaciones *in silico*, sugieren que aquellas proteínas con dirección nuclear en la célula huésped, pueden estar desempeñando un rol importante en la etiología del cáncer. Por ejemplo, entre las proteínas predichas con direccionamiento nuclear, está la subunidad RecB de la exonucleasa V de *C. pneumoniae* o la endoribonucleasa SymE en *S. typhimurium*.

2. PREGUNTA DE INVESTIGACIÓN Y OBJETIVOS

2.1. Descripción del problema

El cáncer somático es originado principalmente por tres factores: la dieta, el estilo de vida y la exposición a agentes carcinogénicos incluidos ciertos agentes infecciosos (7). Aproximadamente un 20% de las neoplasias se desarrollan debido a infecciones crónicas causadas por parásitos, virus o bacterias (4). La bacteria *H. pylori* es, de momento, la única reconocida como agente carcinogénico, siendo un factor etiológico del cáncer de estómago. En este sentido, hay particular interés en conocer qué otros agentes bacterianos pueden causar cáncer (34-39). Así, mediante el reporte de casos y estudios epidemiológicos, se ha sugerido que *S. typhi* y *M. tuberculosis* se asocian con el cáncer de vesícula y de pulmón, respectivamente, aunque la causalidad aún no ha sido demostrada (4,18,20). Puntualmente, la asociación entre la infección por tuberculosis y el CP está bien documentada a través de varios estudios, aunque los mecanismos moleculares detrás de dicha asociación aún no han sido dilucidados. (20-22).

Sumado a la evidencia epidemiológica que relaciona la infección por tuberculosis y el CP, algunos estudios *in vitro* han abordado esta problemática. Así, se ha sugerido que *M. tuberculosis* podría inducir el CP a través de alguno de los mecanismos descritos en organismos patógenos carcinogénicos. Una posibilidad es que promueva la transformación de manera indirecta, por el ambiente intracelular generado por la inflamación crónica en el tejido

pulmonar. Ello ocurre en los pacientes con *M. tuberculosis* activa, donde los granulomas generan una respuesta inflamatoria que daña los tejidos, los cuales son reparados incrementando la tasa de proliferación celular. Esta última se asocia con errores en la replicación cromosómica, que finalmente conllevan a mutaciones en el ADN. Sumado al proceso de angiogénesis que también es crucial en la reparación tisular, la tuberculosis crónica podría ser un caldo de cultivo para la formación de tumores (18,19). También, se ha planteado la posibilidad de la transferencia génica debido a que *M. tuberculosis* es una bacteria intracelular (18). Es decir, la micobacteria podría integrar su ADN plasmídico en el genoma humano y así promover la carcinogénesis pulmonar. Otra hipótesis se relaciona con las mutaciones del gen de la tríada frágil de la histidina, que en pacientes con tuberculosis pulmonar crónica, induciría la carcinogénesis pulmonar (20). Otro posible mecanismo es la acción de proteínas con capacidad genotóxica liberadas por *M. tuberculosis*, el cual es uno de los mecanismos atribuidos a la carcinogénesis promovida por las bacterias (1,7,8). Las genotoxinas son proteínas tóxicas, secretadas por las bacterias, y dirigidas al núcleo que dañan la integridad del ADN de simple o doble cadena, provocando, por ejemplo, cortes en la cadena del ácido nucleico tal, como se ha descrito en bacterias intestinales (1,7,8). Dichos cambios en el genoma pueden afectar la función de los protooncogenes y genes supresores de tumor, lo cual, a la larga, promovería la carcinogénesis (7). Ello ha sido descrito en *S. typhi*, una bacteria gram-negativa que secreta la genotoxina CdtB que provoca mutaciones en p53 e inducción del cáncer de vías biliares (40). En el caso de *M. tuberculosis*, se ha descrito mediante ensayos *in vitro* que puede ocasionar alteraciones en el ADN, lo cual podría ser un desencadenante de la transformación (23,24).

La micobacteria *M. tuberculosis*, al igual que otras bacterias infecciosas, tiene factores de virulencia que son críticos para la patogénesis de la infección. Si alguno de esos factores muestra una actividad genotóxica y desencadena la transformación, es algo aún desconocido. Sin embargo, es plausible determinar la presencia de proteínas secretoras con afinidad por el ADN, analizando datos del proteoma y del secretoma de la cepa *H37Rv* de *M. tuberculosis*. Tanto el proteoma como el secretoma de la micobacteria han sido determinados experimentalmente y han permitido la identificación de algunos polipéptidos que median la interacción huésped-patógeno, los cuales constituyen potenciales dianas terapéuticas y candidatos a vacunas (25-28). Sin embargo, no se sabe aún si parte del proteoma o secretoma de la micobacteria es destinado al interior de la célula huésped ni tampoco el destino subcelular de dichas proteínas de la micobacteria internalizadas a la célula eucariota. Así mismo, es una

incógnita si alguna proteína de *M. tuberculosis* es destinada al núcleo de la célula humana y si dicho polipéptido podría asociarse con el ADN.

Dado que las secuencias del proteoma y secretoma de *M. tuberculosis* están disponibles en bases de datos públicas (27-33), determinar las características de las proteínas de interés y su localización subcelular es viable. Por ende, el acceso al proteoma y secretoma completo de *M. tuberculosis* es viable, rápido y sencillo por lo que su análisis secundario es plausible.

2.2. Justificación

A nivel mundial, se diagnosticaron alrededor de 2,2 millones de casos de CP en el 2020, siendo el segundo tipo de cáncer más frecuente y el primero más mortal (41). Si bien hasta dos tercios de casos de CP son originados por el consumo de tabaco, otros factores como la polución y enfermedades crónicas previas, como COPD o infección por tuberculosis también estarían contribuyendo a la neoplasia (19,41).

Por otro lado, la incidencia anual de tuberculosis pulmonar es de 10 por cada 100000 personas, en países de ingresos altos, y de 183 por 100000 personas en países con ingresos bajos y medios (15). La tuberculosis es una enfermedad prevalente en los países en desarrollo o subdesarrollados, especialmente en regiones donde predomina la pobreza, como en África. En la región de las Américas, el Perú es el segundo país con más casos de tuberculosis reportados, después de Brasil. En nuestro país, se estima que se diagnosticaron 27 000 casos de tuberculosis en el 2022 (42).

Las personas con tuberculosis pulmonar previa tienen un mayor riesgo de desarrollar CP que las personas sin tuberculosis, independientemente del hábito de consumo de tabaco (18-22). Además, lesiones en los sitios de infección previa por *M. tuberculosis* en pulmón son lugares donde se desarrollan carcinomas (20). Todo ello sugiere que la micobacteria podría contribuir al desarrollo del CP, aunque ello aún no ha sido comprobado ni mucho menos se ha determinado un mecanismo mediador. Por lo tanto, dado el interés que suscita la relación entre dos enfermedades de gran interés en salud pública, es razonable investigar los posibles mecanismos mediadores del cáncer, desplegados por *M. tuberculosis* en el epitelio pulmonar.

M. tuberculosis no solo causa anomalías nucleares *in vitro*, también ocasiona cortes en el ADN de macrófagos, activa la señalización de ATM-AKT e inhibe la apoptosis (23,24). Dichos efectos serían mediados por un grupo de factores de la micobacteria, los cuales aún no han sido determinados. Por lo expuesto, conocer si la micobacteria libera ciertos factores que dañen el

material genético, de una manera directa a través de su interacción con el ADN, sería una evidencia importante sobre la interacción patógeno-huésped a nivel del núcleo eucariota. Teniendo como antecedente el hecho de que ciertas bacterias intracelulares vinculadas al cáncer secretan factores que se dirigen al núcleo de la célula huésped, el presente estudio propone llenar un vacío de conocimiento en torno a *M. tuberculosis* (una bacteria intracelular facultativa) y a sus factores proteicos que potencialmente pueden alterar el ADN (34-39).

Este conocimiento podría cambiar nuestra visión acerca de la micobacteria causante de la tuberculosis y su interacción con el huésped. En un futuro, se podría comprobar la interacción entre proteínas de la micobacteria con el material genético humano mediante experimentos dirigidos, pudiendo así dilucidar su potencial capacidad genotóxica en las células huésped. Los resultados de estas investigaciones podrían impactar en las políticas de salud pública de prevención y manejo de la tuberculosis que no solo consideren a la micobacteria como un agente infeccioso sino también como un agente genotóxico.

2.3. Pregunta de investigación

¿El proteoma de *M. tuberculosis H37Rv* contiene proteínas de secreción que se dirigen al núcleo de las células humanas e interactúan con el ADN?

2.4. Objetivo general

Predecir las proteínas de secreción de *M. tuberculosis H37Rv* que potencialmente se dirijan al núcleo de las células humanas e interactúen con el ADN.

2.5. Objetivos específicos

- Predecir las proteínas de secreción de *M. tuberculosis H37Rv* y su direccionamiento nuclear en las células humanas.
- Identificar proteínas de secreción de *M. tuberculosis H37Rv* dirigidas al núcleo que sean homólogas con genotoxinas bacterianas.
- Mapear regiones y/o dominios de unión al ADN en las proteínas de secreción de *M. tuberculosis H37Rv* dirigidas al núcleo.
- Predecir los complejos proteína-ADN de proteínas de secreción de *M. tuberculosis H37Rv* dirigidas al núcleo.

3. MATERIALES Y MÉTODOS

3.1. Diseño del estudio.

El presente estudio es íntegramente teórico-computacional, prospectivo, de análisis secundario de datos ómicos depositados en recursos públicos. En la Fig. 1 se presenta el algoritmo de la investigación.

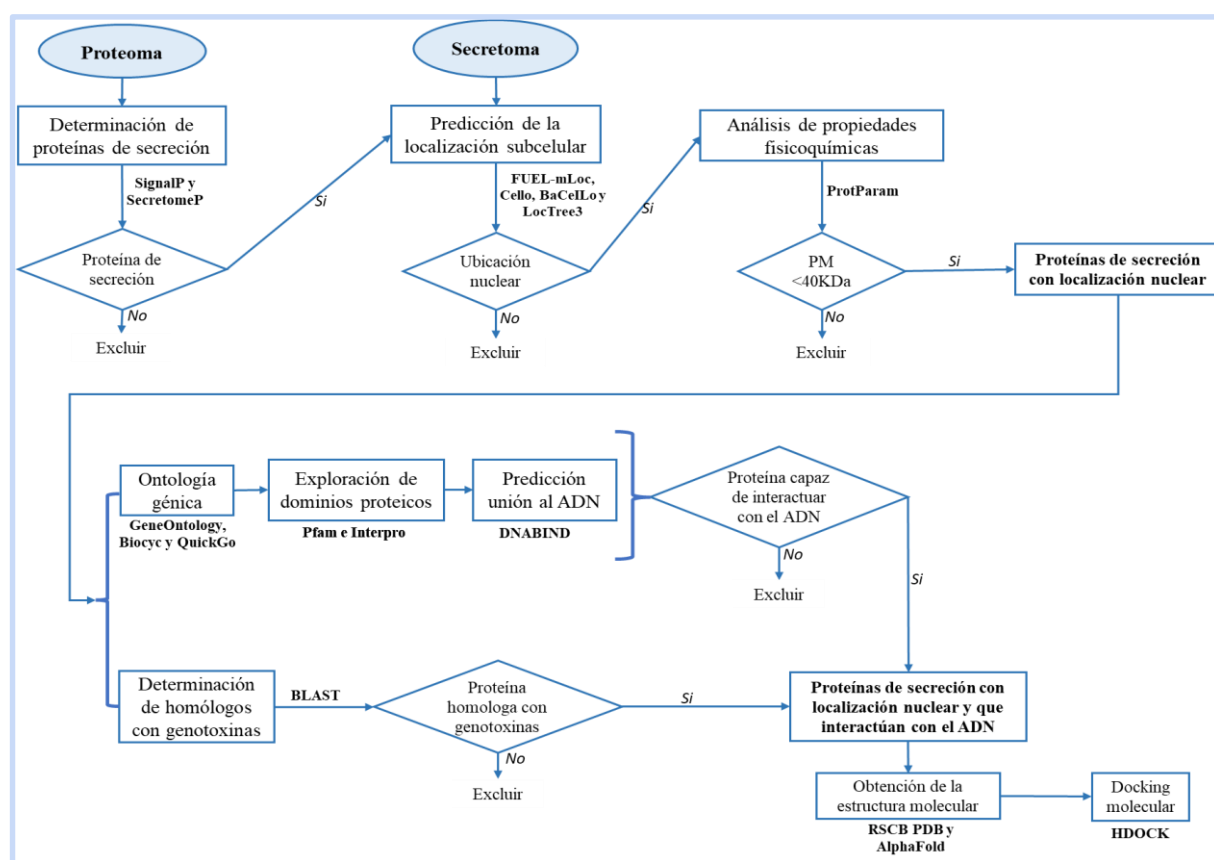


Figura 1. Flujograma de trabajo.

3.2. Obtención de la base de datos del proteoma y secretoma de *M. tuberculosis*.

Se realizó una búsqueda bibliográfica de estudios que habían determinado el proteoma de la cepa bacteriana *M. tuberculosis H37Rv*, consultando motores de búsqueda como PubMed, ResearchGate y Google académico. Se recuperó la información de las proteínas identificadas en dichas publicaciones y se unificaron en una base de datos *in house* correspondiente al proteoma unificado. Las proteínas repetidas entre un artículo y otro fueron consideradas sólo una vez y reportadas como proteínas comunes entre al menos dos artículos.

De igual modo, usando los motores de búsqueda mencionados, se buscaron trabajos que habían determinado experimentalmente el secretoma de *M. tuberculosis H37Rv*. La información de dichas proteínas se vació en una base de datos unificada del secretoma. Para todas las proteínas, se incluyó la siguiente información: el uniprot ID, el nombre de gen, el tamaño en aminoácidos, la técnica usada en su determinación y la referencia bibliográfica. Para el secretoma experimental, se recuperó de cada publicación el número Gi de las proteínas, el cual fue convertido al código de Uniprot ID utilizando la herramienta ID Mapping de Uniprot (43). Esta conversión fue hecha para uniformizar el formato de identificador de todas las proteínas.

Para las pruebas subsecuentes, se descargó en formato fasta la secuencia de las proteínas de estudio, tanto del grupo de proteoma como del secretoma. Ello se realizó ingresando la lista de todos los Uniprot IDs de las proteínas de interés en Uniprot, seleccionando a *Mycobacterium tuberculosis* (strain ATCC 25618 / H37Rv) como organismo de interés. También se usó el GenBank como un buscador complementario para obtener las proteínas del secretoma experimental, en los casos en el que la proteína de interés no se encuentre en la base de datos de Uniprot (44).

3.3. Determinación *in silico* de proteínas de secreción.

Se predijeron las proteínas de secreción aplicando métodos teóricos. Para ello, primero se ingresaron a SignalP-5.0 las secuencias de aminoácidos de cada polipéptido del proteoma unificado de la micobacteria. El programa SignalP determinó la probabilidad de que dichos polipéptidos fueran de secreción, considerando la presencia de un péptido señal (vía clásica de secreción) (45). Luego, aquellas proteínas no predichas de secreción por SignalP-5.0 fueron ingresadas al programa SecretomeP, el cual predice las proteínas que son secretadas por un mecanismo no basado en un péptido señal (vía no clásica de secreción) (46). De este modo, se obtuvo una lista única de proteínas predichas de secreción, con presencia o ausencia de un péptido señal. Este grupo de proteínas se denominó “secretoma teórico” el cual fue comparado con el secretoma experimental usando el programa Venny (<https://bioinformatics.psb.ugent.be/webtools/Venn/>) y valiéndonos de los Uniprot IDs.

3.4. Predicción de la localización subcelular en las células huésped.

Tanto aquellas proteínas de *M. tuberculosis* predichas de secreción en nuestro estudio, así como aquellas determinadas experimentalmente de secreción (secretoma), fueron analizadas

mediante los programas FUEL-mLoc, Cello, BaCelLo y LocTree3, con el fin de conocer su localización subcelular y determinar si son destinadas al núcleo (47-50).

Para ello, se ingresó las secuencias de las proteínas a cada programa y se seleccionó el tipo de célula de destino. En el programa FUEL-mLoc, se predijo la localización subcelular de los polipéptidos simulando el destino en células humanas, el cual predice hasta 14 sub-localizaciones (47). En tanto, la prueba con BaCelLo se llevó a cabo considerando el destino en células eucariotas animales, que analiza 4 sub-localizaciones (48). Finalmente, en los programas Cello y LocTree3 se condujo el análisis de destino en células eucariotas, incluyendo hasta 12 y 19 sub-localizaciones celulares, respectivamente (49,50). Cabe mencionar que en LocTree3, las proteínas que resultaron con una sub-localización en el plastidio o cloroplasto fueron reportadas de localización mitocondrial, de acuerdo a las pautas del programa (50). Los resultados obtenidos fueron tabulados para cada polipéptido de tal forma que se consideraron proteínas destinadas al núcleo aquellas que contaban con un resultado positivo para la localización nuclear, en por lo menos 2 de los 4 programas empleados.

3.5. Análisis de las propiedades fisicoquímicas de las proteínas de localización nuclear.

Usando el programa ProtParam, se determinaron las principales propiedades fisicoquímicas tanto de las proteínas del secretoma y las de secreción predichas, con localización nuclear (51). Para ello, se introdujeron, de una en una, las secuencias de las proteínas en el programa. Esta herramienta computó distintos parámetros fisicoquímicos, de entre los que se eligió el punto isoeléctrico (PI) y el peso molecular (PM) (51). Para los posteriores análisis, se seleccionó únicamente aquellas proteínas con un PM menor a 40 KDa, puesto que son las más probables de difundirse libremente e ingresar al núcleo de las células huésped (14, 52).

3.6. Ontología génica, exploración de dominios proteicos y predicción de unión al ADN

En esta etapa se estudió a las proteínas de secreción (del secretoma experimental y teórico) de *M. tuberculosis* con un PM < 40 KDa y destinadas al núcleo. En primer lugar, se realizó un análisis de su ontología génica en los programas GeneOntology, Biocyc y QuickGo (53-55). Estas herramientas estimaron los 3 componentes de la ontología génica de cada proteína, incluyendo la función molecular, su rol en procesos biológicos y el componente celular. El análisis se realizó, principalmente en GeneOntology, ingresando el código Uniprot (Uniprot ID) de las proteínas de interés (53). Un grupo de proteínas en las que no se obtuvo resultados

con GeneOntology, fueron analizadas con Biocyc y Quickgo (54,55). Además, cuando alguno de los programas no pudo reconocer el código Uniprot de una proteína, se ingresó el nombre del gen al programa.

A continuación, el grupo de proteínas ya mencionadas, se sometió a una exploración de dominios proteicos. Para ello, se ingresaron las secuencias de cada proteína en los programas Pfam e Interpro y luego se identificó, en particular, dominios involucrados en la unión al ADN (56, 57). Por último, se exploró con el programa DNABIND, la capacidad de unión al ADN de las proteínas de nuestro interés (58). Para esta prueba, se ingresaron a DNABIND las secuencias de las proteínas. Este programa usa un algoritmo de predicción basado en plantillas y en un aprendizaje automático, para reconocer o inferir sitios de unión al ADN (58). Los resultados obtenidos, se tabularon en hojas de Excel, por cada proteína y cada estudio realizado. Para los siguientes pasos, solo se seleccionaron las proteínas que en los tres análisis (ontología génica, dominio proteico y predicción de unión al ADN) mostraban indicios de una posible unión al ADN.

3.7. Determinación de homología con genotoxinas bacterianas

Para averiguar si las proteínas del secretoma y las predichas de secreción de *M. tuberculosis*, dirigidas al núcleo y con PM menor a 40 KDa, eran homólogas de genotoxinas bacterianas, se realizó un análisis con BLAST (59). La exploración se inició seleccionando una genotoxina clásica, cuya capacidad de dañar el ADN está bien estudiada. Se eligió la CDT (toxina de distensión citoletal), específicamente la subunidad B (CdtB) (8,10). Para ello, se procedió a buscar su secuencia en Uniprot y se descargó aleatoriamente la secuencia de tres bacterias: *E. coli* (ID Uniprot: Q46669), *S. typhi* (ID Uniprot: Q8Z6A7) y *Campylobacter hyointestinalis* (ID Uniprot: A0A024FRP4). Luego se ingresó las secuencias seleccionadas en el BLAST y se escogió el proteoma de *M. tuberculosis H37Rv* (Taxid: 83332) como fuente de búsqueda en el BLAST.

3.8. Modelamiento y visualización molecular

En esta etapa, se buscó en el Protein Data Bank (RSCB PDB) la estructura tridimensional (3D) de las proteínas de *M. tuberculosis* con capacidad de unirse al ADN, según lo analizado en la sección 3.6. Toda vez que la estructura 3D estaba disponible, la misma se descargó en formato pdb para su posterior visualización. Para las proteínas cuyas estructuras 3D no estaban disponibles en el PDB, los modelos fueron obtenidos de AlphaFold, una base de datos de

estructuras 3D modeladas usando algoritmos de *machine learning* (61). Para ello, se ingresó el código Uniprot en el buscador de AlphaFold.

Una vez identificadas las estructuras de interés en AlphaFold, las mismas fueron descargadas en formato pdb y luego cargadas en el servidor DeepRefiner para ser refinadas energéticamente, dejando los parámetros del programa por defecto (62). Este servidor, usa el aprendizaje profundo para mejorar la precisión de una estructura y acercar su calidad a un nivel experimental, proporcionando modelos en base a parámetros como la puntuación de calidad global, de energía de Rosseta, MolProbity, GOAP, OPUS-PSP, DFIRE y RWplus (62). El modelo refinado elegido, en base a la puntuación de calidad global, se visualizó y comparó con la estructura no refinada para calcular el rmsd, usando el programa ChimeraX (63), con el fin de verificar la *performance* del refinamiento.

3.9. Docking molecular

Por último, con la finalidad de modelar los complejos entre las proteínas de interés y el ADN/ARN, se llevó a cabo una simulación de acoplamiento molecular en el programa HDOCK (64). Para ello, primero se buscó y recuperó, del RSCB PDB, estructuras 3D de complejos ADN-proteína donde ésta última perteneciera a la misma familia que el grupo de proteínas de estudio de esta tesis o bien que presentara los mismos dominios de unión al ADN que las proteínas de la micobacteria, según lo descrito en la sección 3.6. La búsqueda de estos complejos se hizo con el fin de conocer el patrón de unión al ADN/ARN de este tipo de proteínas y de contar con fragmentos de ADN/ARN reconocidos por dichos polipéptidos. Usando el programa ChimeraX (63), se extrajeron las estructuras del ADN de doble hebra o bien de ARN de los complejos resueltos, los cuales fueron usados para las simulaciones de acoplamiento con las proteínas de interés de *M. tuberculosis*. Las secuencias de empleadas para el docking fueron: S1: 5'-CGGAACATGTAAATAGTTACATGATTAC-3', ADN de doble cadena, perteneciente a *Bacillus subtilis*; S2: 5'-GGCTTCCTCCTGG-3', ADN de doble cadena que corresponde a una secuencia sintetizada de *E.coli*; y S3: 5'-CCpCGCGGAUCAGUCACCCAAGCGAG-3', hebra simple de ARN de humano.

Luego, se procedió a correr la simulación de acoplamiento, ingresando la estructura de la proteína (receptor) y al ADN/ARN (ligando) al programa HDOCK, ambos en formato PDB. Se simularon tres complejos: (i) la estructura de la proteína resuelta I6X7F9 (PDB: 5ZHC) con la secuencia S1; (ii) la estructura modelada y refinada de P9WNC3 con S2 y (iii) la estructura resuelta de la proteína P9WH37 (PDB:7 KGB) con S3. A continuación se inició la simulación

bajo los parámetros por defecto y sin adicionar criterios en la opción avanzada, como la selección de residuos del sitio de unión en las proteínas de estudio. Una vez finalizada la simulación, se obtuvieron las mejores conformaciones del complejo proteína-ADN/ARN, cada uno con sus respectivos valores de la puntuación de acoplamiento (*docking scoring*). Se eligió la conformación de menor valor de puntuación de acoplamiento (valor más negativo), la cual se procedió a visualizar y analizar con el programa ChimeraX, identificando los residuos interactuantes (63). Asimismo, se usó PLIP (Perfilador de interacción proteína-ligando) para identificar las interacciones proteína-ADN/ARN en los complejos modelados (65). Para ello, se ingresó cada complejo modelado y se corrió el análisis con los parámetros por defecto, considerando a los ácidos nucleicos como ligandos. Los resultados fueron visualizados y luego descargados en ficheros para su posterior análisis.

3.10. Cuestiones éticas

En el presente estudio no hubo sujetos participantes, ni pruebas con animales ni células. Se obtuvo la exoneración de la aprobación por el Comité Institucional de Ética de la UPCH (constancia CAREG-ORVEI-057-22). Los datos utilizados son de libre acceso para el público, por lo que su uso posterior es viable.

4. RESULTADOS

Los resultados obtenidos en los diferentes procesos del presente estudio se pueden observar en la Fig.2, la cual también muestra cómo variaron los conjuntos de datos, así como el número de proteínas analizadas.

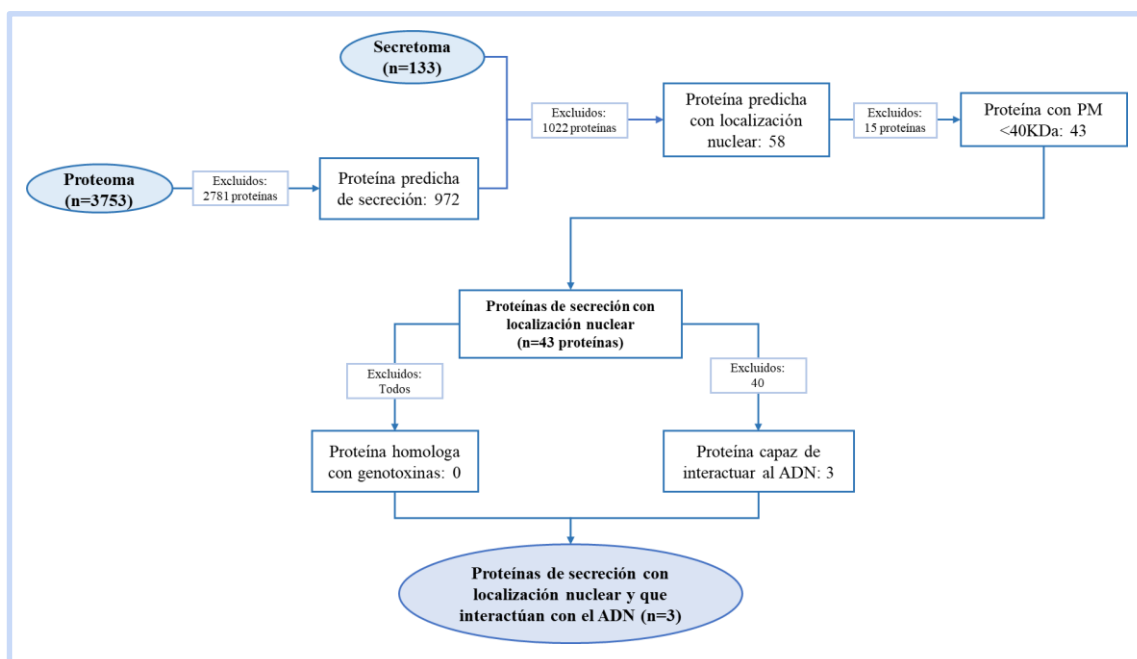


Figura 2. Diagrama de representación de resultados obtenidos.

4.1. Compilación del proteoma de *M. tuberculosis*

La composición del proteoma de *M. tuberculosis H37Rv* ha sido reportada en 6 estudios experimentales, los cuales emplearon la técnica de espectrometría de masas acoplada a cromatografía líquida (LC-MS/MS), según se muestra en la Tabla 1 (28-33). El número de proteínas identificadas en cada estudio varió entre 1051 y 4151.

Tabla 1. Publicaciones del proteoma de *M. tuberculosis* utilizadas en este estudio.

	Tucci et al., 2020	Kelkar et al., 2011	Bell et al., 2012	Peters et al., 2016	Gopinath et al., 2015	Albrethsen et al., 2013
Número de proteínas reportadas	1314	3176	1051	3788	1871	1176
Referencia bibliográfica	(27)	(28)	(29)	(30)	(31)	(32)

A partir de la información contenida en dichas publicaciones, se obtuvo una lista unificada del proteoma de la micobacteria, conformada por 3753 proteínas (Anexo 1). Basándonos en el nombre de la proteína declarado en dichas publicaciones, se observó que 715 polipéptidos fueron funcionalmente “no caracterizados”, mientras que 589 tenían una predicción funcional “probable”, 280 eran proteínas anotadas con funciones “posibles”, 208 eran proteínas con

funciones putativas y 178 eran proteínas conservadas y sin función anotada. El resto de las 1783 proteínas, sí poseían una función reconocida (anexo 1). El tamaño de las proteínas del proteoma de *M. tuberculosis* fluctuó entre 27 y 4151 aminoácidos (anexo 1). Así mismo, identificamos 593 proteínas del proteoma que fueron comunes en los seis estudios considerados.

Por otro lado, se identificó sólo un estudio donde determinaron el secretoma de *M. tuberculosis* (34). En dicha investigación, mediante la técnica de *shotgun* proteómica (*shotgun-proteomics*), se reveló que el secretoma de *M. tuberculosis H37Rv* estaba compuesto por 837 proteínas (34). Al ingresar los códigos de dichos polipéptidos en Uniprot, solo se logró recuperar un total de 133 polipéptidos, mientras que en 704 proteínas no hubo alguna secuencia identificada en *M. tuberculosis H37Rv* o bien el resultado indicaba que pertenecían a otros microorganismos o cepa micobacteriana o simplemente sus números de GI y Uniprot ID no pudieron ser reconocidos. Puesto que el número de proteínas recuperadas fue muy bajo, se consultó también el GenBank, pero los resultados fueron los mismos. Por ello, sólo fueron considerados los 133 polipéptidos en los análisis subsecuentes (Anexo 2). Dentro de este grupo, 20 proteínas resultaron ser “no caracterizadas” funcionalmente, 14 polipéptidos estaban anotados con una función “probable”, 11 proteínas con funciones “posibles”, 5 polipéptidos con funciones putativas, 5 eran proteínas conservadas y 77 presentaban una función validada (Anexo 2). Asimismo, el tamaño de las proteínas del secretoma de *M. tuberculosis* fluctuó entre 53 y 1205 aminoácidos (Anexo 2).

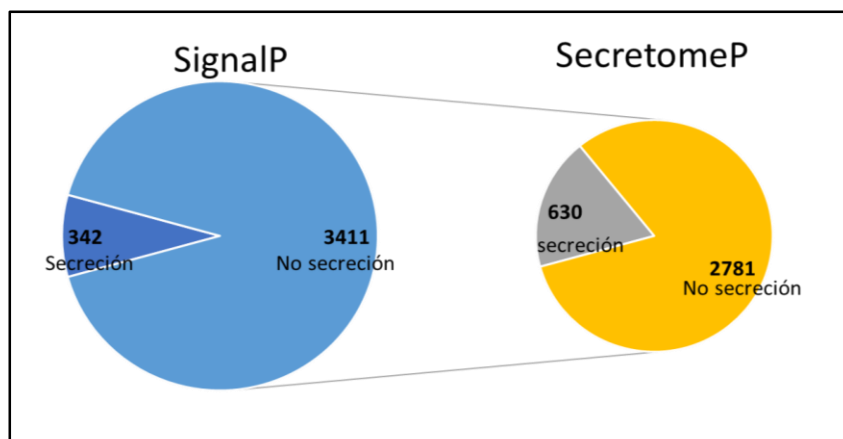
4.2. Predicción de las proteínas de secreción

Con el fin de predecir las proteínas de secreción que formaban parte del proteoma de *M. tuberculosis*, se emplearon dos programas, SignalP y SecretomeP. Primero, se realizó el análisis de los 3753 polipéptidos del proteoma en el programa SignalP. Se logró identificar 342 proteínas, las cuales presentaban una alta probabilidad de tener un péptido señal (PS), es decir, fueron predichas de secreción (Fig. 3, anexo 3). Entre los péptidos señales identificados se incluye el de lipoproteína (Sec/SPII), Sec (Sec/SPI), y Tat (Tat/SPI) (Anexo 3).

Luego, se analizaron con SecretomeP las 3411 proteínas en las que no se encontró algún PS. De esta prueba, se detectaron 630 proteínas que mostraron una alta probabilidad (SecP score > 0.5) de ser secretadas por la vía secretora no clásica, es decir, no mediadas por PS (Fig. 3, anexo 4). En total, 972 polipéptidos de *M. tuberculosis* fueron predichos de secreción, incluyendo

proteínas con péptidos señal o sin éste (Anexos 3 y 4). A este grupo se le denominó “secretoma teórico”.

Figura 3. Predicción de proteínas de secreción de *M. tuberculosis H37Rv* mediante los programas SignalP (vía clásica) y SecretomeP (vía no clásica).



Por otro lado, con el interés de estimar la precisión del algoritmo de predicción de secreción, se comparó el secretoma teórico con el secretoma experimental recuperado (n=133), empleando el programa Venny. De las 972 proteínas predichas de secreción, 25 polipéptidos pertenecían al secretoma experimental (Tabla 2). Además, se compararon las 3753 proteínas del proteoma con estas 133 del secretoma recuperado y resultaron 127 proteínas comunes mientras que otras 6 proteínas del secretoma no estaban contenidas en el proteoma incluyendo I6XFT2, I6X642, I6X8G2, A0A089QKZ7, O53173, B5TV88.

Tabla 2. Polipéptidos en común entre el secretoma teórico (n=972) y el secretoma experimental recuperado (n=133).

Uniprot ID	Nombre de la proteína	Uniprot ID	Nombre de la proteína
O53734	Proteína que contiene el dominio amino_oxidasa	P9WGU3	Proteína de unión a molibdato ModA
P9WN33	Putativa asparagina sintetasa [hidrolizante de glutamina]	P9WQP3	Diacilglicerol aciltransferasa / micoliltransferasa Ag85A
P9WKU3	Proteína no caracterizada Rv0497	P9WP69	Subunidad 2 de citocromo c oxidasa
P9WP37	Citocromo bc1 complejo citocromo b subunidad	P9WPG3	Probable fosfatidilglicerofosfato sintasa
P9WJQ1	Proteína similar a nitrato reductasa NarX	P9WK81	Lipoproteína putativa LppA
I6YC99	Proteína de la familia Mce Mce4A	P9WJB3	Proteína permeasa de exportación de fluoroquinolonas Rv2686c

I6YGV9	Probable proteína de membrana integral transportadora de dipéptidos Transportador ABC DppB	P9WFZ9	Supuesta proteína permeasa transportadora de péptidos Rv1282c
I6YAY5	Proteína no caracterizada	O05896	Posible proteína de membrana integral de transporte catiónico de aminoácidos
O06312	Transportador de magnesio MgtE	I6YGX2	Peptidoglicano glicosiltransferasa
P9WNK5	Proteína similar a ESAT-6 EsxB	P9WFS3	Proteína de membrana de la familia TVP38/TMEM64 Rv1491c
P9WNM3	Factor de elongación P	P71734	Proteína conservada (sin nombre función reconocida)
P9WK09	Metalotioneína	P71883	Probable proteína de extrusión de nitrito 1 NarK1 (facilitador de nitrito 1)
P9WHR1	Supuesta metaloproteasa de zinc Rip3		

4.3. Localización subcelular de las proteínas de secreción.

Con el fin de identificar las proteínas de secreción de *M. tuberculosis H37Rv* que se dirigen al núcleo, tanto el secretoma teórico como las 133 proteínas recuperadas del secretoma experimental fueron analizadas en 4 programas (FUEL-mLoc, Cello, BaCeILo y LocTree3). La localización subcelular fue obtenida primero para el secretoma teórico y luego para el experimental. Cabe mencionar que las 25 proteínas de secreción en común (Tabla 1), fueron consideradas dentro del grupo de secretoma teórico, por lo que un total de 108 proteínas formaron parte del análisis del grupo del secretoma experimental.

Con el programa FUEL-mLoc, se predijo que 57 de las 972 proteínas (5.9%) de secreción predichas y 4 de las 108 (3.7%) proteínas del secretoma experimental tenían como destino el núcleo (Tablas 3 y 4). En tanto, con el programa Cello, se predijo que 149 de las 972 proteínas predichas de secreción (15.3%) y 18 de las 108 proteínas experimentales (16.7%), eran teóricamente dirigidas al núcleo (Tablas 3 y 4). Por su parte, usando BaCeILo, se observó que 30 de las 972 proteínas de secreción teóricas (3.1%) y 9 de las 108 proteínas del secretoma (8.3%) tenían como destino el núcleo (Tablas 3 y 4). Por último, en el programa LocTree3, 37 de los 972 polipéptidos de secreción teóricos (3.8%) y 9 de las 108 proteínas del secretoma experimental (8.3%), eran teóricamente nucleares (Tablas 3 y 4).

Tabla 3. Análisis de localización subcelular del secretoma teórico (n=972) mediante los programas FUEL-mLoc, Cello, BaCeILo y LocTree3.

FUEL-mLoc	n	Cello	n	BaCelLo	n	LocTree3	n
Membrana plasmática	583	Membrana plasmática	524	Mitocondria	407	Secretado (Extracelular)	255
Citoplasma	130	Extracelular	218	Vía secretora	325	Membrana plasmática	213
Extracelular	131	Núcleo	149	Citoplasma	210	Citoplasma	137
Núcleo	57	Cloroplasto	146	Núcleo	30	Membrana de la mitocondria	97
Sinapsis	31	Mitocondria	117			Mitocondria	75
Mitocondria	26	Citoplasma	84			Membrana del aparato de golgi	67
Retículo endoplasmático	18	Vacuola	2			Membrana del retículo endoplasmático	49
Microsoma	8	Peroxisoma	2			Núcleo	37
Citoesqueleto	6	Lisosoma	1			Membrana de la vacuola	21
Lisosoma	3	Retículo endoplasmático	0			No identificado	7
Aparato de Golgi	5	Citoesqueleto	0			Peroxisoma	5
Peroxisoma	4	Golgi	0			Retículo endoplasmático	3
Endosoma	1					Vacuola	3
Centrosoma	1					Membrana del núcleo	2
						Membrana del peroxisoma	1
						Plastidio	0
						Cloroplasto	0
						Membrana del cloroplasto	0
						Aparato de Golgi	0

Tabla 4. Análisis de localización subcelular de las proteínas del secretoma experimental (n=108), mediante los programas FUEL-mLoc, Cello, BaCelLo y LocTree3.

FUEL-mLoc	n	Cello	n	BaCelLo	n	LocTree3	n
Membrana plasmática	39	Citoplasma	59	Citoplasma	57	Citoplasma	41
Citoplasma	37	Mitocondria	46	Mitocondria	36	Mitocondria	40
Extracelular	12	Cloroplasto	32	Núcleo	9	Extracelular	10
Mitocondria	8	Núcleo	18	Vía secretora	6	Núcleo	9
Núcleo	4	Membrana plasmática	9			Peroxisoma	4
Sinapsis	4	Peroxisoma	4			Membrana plasmática	1
Microsoma	3	Extracelular	3			Membrana del aparato de Golgi	1
Lisosoma	1	Citoesqueleto	1			Membrana del retículo endoplasmático	1
Retículo	1	Vacuola	0			No identificado	1

endoplasmático					
Endosoma	1	Retículo endoplasmático	0	Membrana de la mitocondria	0
Aparato de Golgi	1	Lisosoma	0	Vacuola	0
Peroxisoma	1	Golgi	0	Retículo endoplasmático	0
Citoesqueleto	0			Membrana del peroxisoma	0
Centrosoma	0			Membrana de la vacuola	0
				Membrana del núcleo	0
				Plastidio	0
				Cloroplasto	0
				Membrana del cloroplasto	0
				Aparato de golgi	0

En este estudio impusimos la condición de que las proteínas debían presentar una predicción de localización nuclear en por lo menos 2 de los 4 programas usados, para considerarlos factores de secreción dirigidos al núcleo. Así, se determinó que 48 de las 972 proteínas del secretoma teórico (4.9%), y 10 de las 108 proteínas del secretoma experimental (9.3%) tendrían como destino el núcleo de la célula huésped (Anexo 5). Sólo las proteínas de secreción predichas como nucleares en esta etapa (n=58), fueron consideradas en los subsecuentes análisis.

4.4. Propiedades fisicoquímicas: Peso molecular (PM) y punto isoelectrico (PI)

Con el programa ProtParam, se determinó el PM y PI de las 48 proteínas del secretoma teórico predichas nucleares. Resultaron un total 36 polipéptidos con PM<40 KDa y 12 proteínas con PM>40 KDa (Fig. 4A, anexo 5). En tanto, de las 10 proteínas del secretoma experimental predichas nucleares, 7 presentaban un PM<40 KDa y 3 un PM> 40 KDa (Fig. 4B, anexo 5).

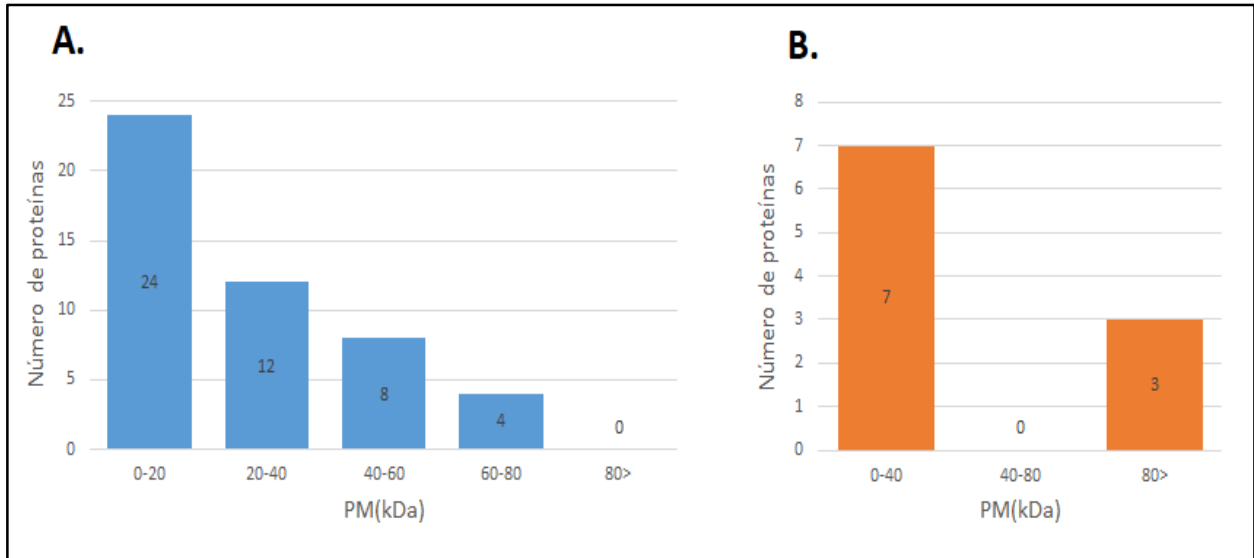


Figura 4. Distribución por peso molecular (PM) de proteínas de secreción de *M. tuberculosis* cuyo destino es el núcleo de la célula huésped. **A:** Proteínas del secretoma teórico. **B:** Proteínas del secretoma experimental.

Por otro lado, las proteínas de secreción de *M. tuberculosis* destinadas al núcleo tuvieron un PI entre 3 a 12 en el grupo del secretoma teórico y entre 3 y 13 en el grupo del secretoma experimental nuclear (Fig. 5A y 5B, anexo 5).

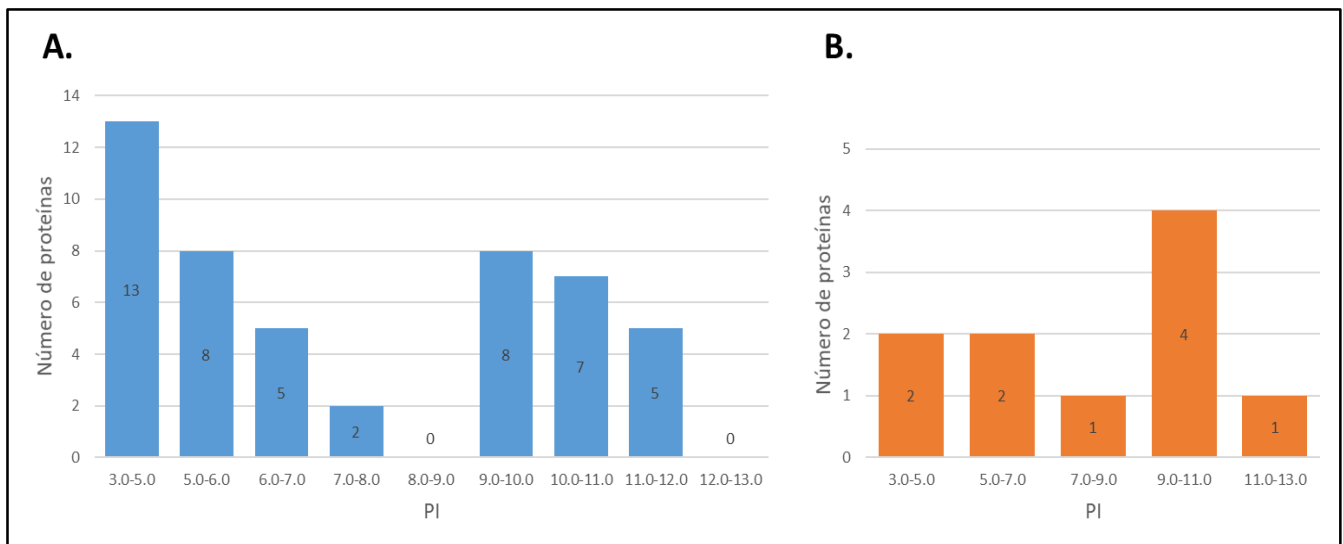


Figura 5. Distribución por punto isoeléctrico (PI) de proteínas de secreción de *M. tuberculosis* cuyo destino es el núcleo de la célula huésped. **A:** Proteínas del secretoma teórico. **B:** Proteínas del secretoma experimental.

Para los posteriores análisis, sólo las proteínas con PM<40 KDa fueron consideradas para las pruebas subsecuentes (n=43), debido a que su probabilidad de ingresar libremente al núcleo es

mayor que aquellas con PM>40 KDa (Tabla 5) (14, 51). De ellas, 23 tuvieron un pI >8 que sería una característica compatible con las proteínas nucleares (66).

Tabla 5. Proteínas de secreción de *M. tuberculosis*, predichas con destino nuclear y de PM<40 KDa.

Uniprot ID	Nombre de la proteína	PM (KDa)	pI
Proteínas del secretoma teórico predichas nucleares			
P9WH95	Proteína ribosómica 50S L33 2	6.4	9.8
I6YFL7	rubredoxina	6.7	3.9
I6XWB9	Proteína que contiene el dominio DUF3073	6.9	5.2
O53937	Probable ferredoxina	7.3	4.1
I6X562	Proteína que contiene el dominio DUF3039	8.9	6.7
O53222	Toxina similar a HicA	9.56	10.2
O53401	Proteína no caracterizada	10.3	8.7
P9WNK5	Proteína similar a ESAT-6 EsxB	10.7	4.5
O05447	Proteína no caracterizada	10.8	4
P9WM27	Proteína no caracterizada Rv1322	11.3	10.7
O05574	Proteína rica en serina conservada	11.4	9
P9WLF7	Proteína no caracterizada Rv2271	11.5	4.8
O53402	Integrasa probable	11.8	10.5
I6X7F9	Regulador transcripcional Rv3488	11.9	9.6
I6Y8X8	transposasa putativa	12.3	6.9
I6YGY9	Proteína rica en prolina hipotética conservada	13	10
P95205	Proteína de la familia de repeticiones 13E12 conservada	13.5	9.9
P9WL11	Proteína no caracterizada Rv3073c	13.7	7.9
I6YHB0	Posible proteína similar a histonas Hns	13.8	10.5
O53621	Probable proteína de membrana	14.6	10.8
P9WMT9	Factor de elongación de la transcripción GreA, Factor de escisión de la transcripción GreA	17.8	4.9
L0T9U5	Proteína del fago PhiRv1 probable	17.9	11.5
L0T8G6	Proteína no caracterizada	18.8	11.9
P95225	Antitoxina putativa Rv0268c	19.4	11.1
P9WK79	Lipoproteína putativa LppB	24.1	6.1
P96376	Posible proteína de membrana conservada	24.5	9.8
P9WLI5	Proteína no caracterizada Rv2206	26.2	11.6
O53572	Proteína de membrana probablemente conservada	27.1	5.8
P9WJC3	Proteína asociada a la secreción de ESX-1 EspJ	27.3	4.1
O05883	Metanotiol S-metiltransferasa	28	10
O06170	Posible proteína de membrana rica en prolina conservada	28.5	7.7
P9WIR9	Homólogo de proteína antigénica de 34 kDa	30.2	6.2
P9WIR7	Proteína secretada rica en alanina y prolina Apa	32.7	4.9
P9WIY3	8-oxo-(d)GTP fosfatasa	34.7	9.2
O06555	Proteína rica en ala conservada	36.4	5.4
O06213	Proteína de membrana rica en prolina probablemente conservada	39.6	11.4

Proteínas del secretoma experimental predichas nucleares			
A0A089QKZ7	Proteína no caracterizada Rv1155A	6.7	10.3
P9WKN9	Proteína no caracterizada Rv0942	9.7	10.8
I6X642	Proteína no caracterizada	10	4.6
B5TV88	Proteína similar a ESAT-6	10.7	4.5
P9WFL1	UPF0232 proteína Rv0004	19.9	11.6
P9WH37	Proteína ribosómica S3 30S	30	10.1
P9WNC3	Formamidopirimidina-DNA glicosilasa 1	31.9	9.9

4.5. Ontología génica, dominios proteicos y predicción de unión al ADN

A continuación, se indagó la ontología génica de las proteínas de interés (Tabla 5), mediante GeneOntology. Los resultados mostraron que, dentro del grupo de proteínas del secretoma teórico predichas nucleares y de PM < 40 Kda, hay proteínas cuya función es la de unión al ADN, así como de unión a la ARN polimerasa, unión a iones y proteínas, unión a la matriz extracelular, constituyentes del ribosoma, entre otros (Tabla 6). Así mismo, este grupo incluye proteínas que regulan procesos biológicos como la reparación del ADN, la evasión de la respuesta inmune del huésped, la transcripción, el crecimiento celular, la recombinación del ADN, entre otros. El componente celular mostró que este grupo contiene proteínas de membrana, extracelulares y citosol.

Tabla 6. Ontología génica de las 36 proteínas del secretoma teórico, con destino nuclear y de PM < 40 KDa, mediante los programas GeneOntolgy, Biocyc y Quickgo.

Uniprot ID	Nombre de la proteína	Función molecular	Rol en los procesos biológicos	Componente celular
P9WIR7	Proteína secretada rica en alanina y prolina Apa	Unión a la matriz extracelular.	Inducción por simbionte de la respuesta inmune del huésped, respuesta a las defensas del anfitrión, respuesta al estímulo biótico, respuesta al estímulo externo, respuesta a la inanición, respuesta celular al estrés.	Superficie celular.
P9WIY3	8-oxo- (d) GTP fosfatasa	Actividad pirofosfatasa de 8-oxo-7,8-dihidroguanosina trifosfato, actividad hidrolasa que actúa sobre los anhídridos de ácido.	Reparación del ADN, respuesta celular al estímulo de daño del ADN, respuesta celular al estrés, replicación del ADN.	Citosol*.

O06555	Proteína rica en ala conservada	No clasificado	No clasificado	Región extracelular**.
P9WK79	Lipoproteína putativa LppB	No clasificado	No clasificado	Membrana del plasma*, región extracelular*, Membrana**.
P9WMT9	Factor de elongación de la transcripción GreA	Unión al ADN*, Unión de ARN polimerasa.	Regulación de la transcripción con plantilla de ADN elongación, transcripción con plantilla de ADN alargamiento, respuesta a antibiótico.	Pared celular, membrana del plasma, Citosol**.
P9WNK5	Proteína EsxB similar a ESAT-6	Unión a proteínas*, unión a la superficie de la célula huésped.	Proceso celular de múltiples organismos, secreción de proteínas por el sistema de secreción de tipo VII, transporte de proteínas, transporte de sustancias orgánicas.	Superficie y retículo endoplásmico de la célula huésped, pared celular*, membrana plasmática*.
I6YHB0	Posible proteína similar a histona Hns	No clasificado	No clasificado	Membrana**.
O53572	Proteína de membrana probablemente conservada	No clasificado	No clasificado	Región extracelular**, espacio periplasmático delimitado por la membrana externa*.
O05574	Proteína rica en serina conservada	Chaperona de plegamiento de proteínas*.	Plegamiento de proteínas*.	Región extracelular*.
P9WH95	Proteína ribosómica 50S L33 2	Constituyente estructural del ribosoma.	Traducción.	Citoplasma, citosol*, ribosoma.
P9WM27	Proteína no caracterizada Rv1322	No clasificado	No clasificado	No clasificado
I6X562	Proteína que contiene el dominio DUF3039	No clasificado	No clasificado	No clasificado
P9WIR9	Homólogo de proteína antigénica de 34 kDa	No clasificado	No clasificado	Componente integral de la membrana plasmática.
P9WLI5	Proteína no caracterizada Rv2206	No clasificado	Evasión de la respuesta inmune del huésped.	Membrana**, componente integral de la membrana**.
P9WJC3	Proteína EspJ asociada a secreción	No clasificado	No clasificado	Región extracelular*.

	de ESX-1			
O06170	Posible proteína de membrana rica en prolina conservada	No clasificado	Crecimiento celular*.	Membrana*, membrana plasmática*.
O06213	Probable proteína de membrana rica en prolina conservada	No clasificado	Crecimiento celular*.	Membrana*.
O05883	Metanotiol S-metiltransferasa	Actividad metiltransferasa, actividad transferasa.	Metilación	Membrana**.
P96376	Posible proteína de membrana conservada	No clasificado	Crecimiento*	No clasificado
O53621	Probable proteína de membrana	No clasificado	No clasificado	Membrana*.
I6X7F9	Regulador transcripcional Rv3488	Unión de ADN, unión de iones metálicos*.	No clasificado	No clasificado
P9WLF7	Proteína no caracterizada Rv2271	No clasificado	No clasificado	No clasificado
O53222	Toxina similar a HicA	No clasificado	No clasificado	No clasificado
P95225	Antitoxina putativa Rv0268c	No clasificado	No clasificado	No clasificado
O05447	Proteína no caracterizada	No clasificado	No clasificado	No clasificado
I6YGY9	Proteína rica en prolina hipotética conservada	No clasificado	No clasificado	No clasificado
I6YFL7	Rubredoxina	Actividad de transferencia de electrones, unión de iones de hierro, unión de iones de metal de transición.	Proceso catabólico de alcanos, proceso catabólico celular, proceso catabólico de hidrocarburos, cadena de transporte de electrones, generación de metabolitos precursores y energía.	Citosol*.
I6XWB9	Proteína que contiene el dominio DUF3073	No clasificado	No clasificado	No clasificado
I6Y8X8	Transposasa putativa	No clasificado	No clasificado	No clasificado
O53401	Proteína no caracterizada	No clasificado	No clasificado	No clasificado
L0T8G6	Proteína no caracterizada	No clasificado	No clasificado	No clasificado
L0T9U5	Probable proteína del	No clasificado	No clasificado	No clasificado

	fago PhiRv1			
P9WL11	Proteína no caracterizada Rv3073c	No clasificado	No clasificado	Citosol*.
O53937	Ferredoxina probable	No clasificado	No clasificado	No clasificado
P95205	Proteína conservada de la familia de repeticiones 13E12	No clasificado	No clasificado	No clasificado
O53402	Integrasa probable	Unión de ADN*.	Integración de ADN, Recombinación de ADN*.	No clasificado

Proteínas que se podrían unir al ADN
Proteínas que se podrían unir al ARN

Todos los resultados mostrados fueron obtenidos mediante GeneOntology, a excepción de los siguientes:

* Resultados obtenidos mediante BioCyc

**Resultados obtenidos mediante QuickGO

La ontología del secretoma experimental, muestra que estas proteínas cumplen funciones moleculares que incluyen la actividad de ADN N-glicosilasa, actividad hidrolasa, unión al ADN de doble hebra, unión ARNs y unión a ácidos nucleicos (Tabla 7). Así mismo, regulan procesos biológicos como la reparación de escisión de la base, respuesta celular al estímulo del daño del ADN y la traducción. El componente celular incluyó membrana plasmática y ribosoma en dos proteínas de este grupo (Tabla 7).

Tabla 7. Ontología génica de las 7 proteínas del secretoma experimental, con un PM < 40 KDa y de ubicación nuclear, mediante los programas GeneOntology, Biocyc y Quickgo.

Uniprot ID	Nombre de la proteína	Función molecular	Rol en los procesos biológicos	Componente celular
P9WNC3	Formamidopirimidina-ADN glicosilasa 1	Actividad de ADN N-glicosilasa de 8-oxo-7,8-dihidroguanina, actividad catalítica actuando sobre el ADN, actividad endonucleasa de ADN de clase I sitio apurínico o apirimidínico, unión de ADN dañada, unión al ADN de doble hebra, unión de iones de zinc.	Reparación de la escisión de la base, respuesta celular al estímulo de daño del ADN, respuesta al estrés oxidativo.	Membrana del plasma**.
A0A089QKZ7	Proteína no caracterizada Rv1155A	No clasificado	No clasificado	No clasificado
P9WH37	Proteína ribosómica 30S S3	Unión a ácidos nucleicos*, Unión de ARNm, unión de ARN, Unión de ARNr, constituyente estructural	Traducción.	Subunidad ribosomal pequeña

		del ribosoma.		citoplasmática.
B5TV88	Proteína similar a ESAT-6	No clasificado	No clasificado	No clasificado
P9WKN9	Proteína no caracterizada Rv0942	No clasificado	No clasificado	No clasificado
P9WFL1	UPF0232 proteína Rv0004	No clasificado	No clasificado	No clasificado
I6X642	Proteína no caracterizada	No clasificado	No clasificado	No clasificado

Proteínas que se podrían unir al ADN
Proteínas que se podrían unir al ARN

Todos los resultados mostrados fueron obtenidos mediante GeneOntology, a excepción de los siguientes:

- * Resultados obtenidos mediante BioCyc
- **Resultados obtenidos mediante QuickGO

A continuación, determinamos los dominios de las proteínas de interés. De las 36 proteínas del secretoma teórico, se reconocieron dominios en 19 de ellas, mediante Pfam, y en 16 de ellas, con Interpro (Tabla 8). Por su parte, se reconocieron dominios en 5 de las 7 proteínas de secreción experimental (Tabla 9). Algunas proteínas presentaron dominios que les confieren la capacidad de interactuar con los ácidos nucleicos, incluyendo el dominio de los dedos de zinc, GreA, GreB, PadR, H2TH, DivIC, el dominio de homología K tipo 2 (KH2), DciA y Antitoxin Phd_YefM (Tablas 8 y 9). Por ejemplo, un factor regulador de la transcripción, un factor de la elongación del ADN, una Formamidopirimidina-ADN glicosilasa 1, entre otros, tendrían dominios capaces de interactuar con el ADN. Además, serían secretadas por la micobacteria y se destinarían al núcleo de la célula humana. Así mismo, un polipéptido ribosómico y una posible proteína de membrana, tendrían dominios de unión al ARN.

Tabla 8. Dominios proteicos obtenidos con Pfam e Interpro en las 36 proteínas del secretoma teórico con un PM < 40 KDa y con ubicación nuclear.

Uniprot ID	Nombre de la proteína	Tamaño en AA	Pfam		InterPro	
			Dominio	Inicio-Fin	Dominio	Inicio-fin
P9WIR7	Proteína secretada rica en alanina y prolina Apa	325	FAP (Proteína de unión a fibronectina)	1-298	Accesorio de fibronectina	1-297
P9WIY3	8-oxo- (d) GTP fosfatasa	317	NUDIX (dominio Nudix)	17-146	Dominio de hidrolasa NUDIX	15-148
			His Phos 1 (Superfamilia de histidina fosfatasa,	156-251	Hidrolasa NUDIX	29-63 / 63-78

			clado 1)			
					Superfamilia de histidina fosfatasa, clado-1	155-296
O06555	Proteína rica en ala conservada	371	NO		NO	
P9WK79	Lipoproteína putativa LppB	220	LppA (Lipoproteína confinada a mycobacterium patógenos)	39-195	NO	
P9WMT9	Factor de elongación de la transcripción GreA	164	GreA GreB N (Factor de elongación de la transcripción, N-terminal)	7-76	Factor de elongación de la transcripción, GreA / GreB, N-terminal	8-76
			GreaA GreB (Factor de elongación de la transcripción, C-terminal)	81-161	Factor de elongación de la transcripción, GreA / GreB, C-terminal	82-160
P9WNK5	Proteína EsxB similar a ESAT-6	100	WXG100 (Proteína de 100 residuos con WXG)	3-88	NO	
I6YHB0	Posible proteína similar a histona Hns	134	NO		NO	
O53572	Proteína de membrana probablemente conservada	264	NO		NO	
O05574	Proteína rica en serina conservada	110	Zn-ribbon8 (Dominio de la cinta de zinc)	1-40	FmdB, dominio de cinta de zinc	1-51
P9WH95	Proteína ribosómica 50S L33 2	55	Ribosomal L33 (Proteína ribosómica L33)	9-55	Proteína ribosómica L33	6-55
P9WM27	Proteína no caracterizada Rv1322	98	NO		NO	
I6X562	Proteína no caracterizada	82	DUF3039 (Proteína de función desconocida)	23-79	Proteína de función desconocida DUF3039	24-79
P9WIR9	Homólogo de proteína antigénica de 34 kDa	303	DUF5336 (Familia de función desconocida)	37-159	Proteína de función desconocida DUF5336	37-159
P9WLI5	Proteína no caracterizada Rv2206	236	DUF3043 (Proteína de función desconocida)	48-233	Proteína de función desconocida DUF3043	48-232

P9WJC3	Proteína EspJ asociada a secreción de ESX-1	280	NO		NO	
O06170	Posible proteína de membrana rica en prolina conservada	273	Mycobac memb (Proteína de la membrana de Mycobacterium)	169-270	NO	
O06213	Probable proteína de membrana rica en prolina conservada	384	NO		NO	
O05883	Metanotiol S-metiltransferasa	244	NO		Dominio NnrU	55-189
P96376	Posible proteína de membrana conservada	228	DivIC (Iniciador de la formación del tabique)	86-164	Iniciador de formación de tabiques FtsL / DivIC	86-161
O53621	Probable proteína de membrana	129	NO		NO	
I6X7F9	Regulador transcripcional Rv3488	107	PadR (Regulador transcripcional de la familia PadR)	13-83	Regulador de transcripción PadR, N-terminal	13-83
P9WLF7	Proteína no caracterizada Rv2271	99	DUF2695 (Proteína de función desconocida)	39-91	Proteína de función desconocida DUF2695	39-91
O53222	Toxina similar a HicA	83	NO		NO	86-161
P95225	Antitoxina putativa Rv0268c	169	PhdYefM_antitox (Antitoxina Phd_YefM, sistema toxina-antitoxina tipo II)	88-150	Sistema de toxina-antitoxina tipo II, antitoxina Phd / YefM	90-142
O05447	Proteína no caracterizada	110	NO		NO	
I6YGY9	Proteína rica en prolina hipotética conservada	129	NO		NO	
I6YFL7	Rubredoxina	60	Rubredoxina (Proteína Rubredoxina)	6-52	Dominio similar a la rubredoxina	4-55
I6XWB9	Proteína que contiene el dominio DUF3073	60	DUF3073 (Proteína de función desconocida)	2-60	Proteína de función desconocida DUF3073	2-58
I6Y8X8	Transposasa putativa	110	NO		NO	
O53401	Proteína no	91	NO		NO	

	caracterizada					
L0T8G6	Proteína no caracterizada	174	NO		NO	
L0T9U5	Probable proteína del fago PhiRv1	166	NO		NO	
P9WL11	Proteína no caracterizada Rv3073c	118	DUF488 (Proteína de función desconocida)	8-115	Proteína de función desconocida DUF488	9-115
O53937	Ferredoxina probable	67	Fer4_15 (Dominio de clúster único 4Fe-4S)	3-61	NO	
P95205	Proteína conservada de la familia de repeticiones 13E12	122	NO		NO	
O53402	Integrasa probable	104	NO		NO	1-104

	Proteínas que se podrían unir al ADN
	Proteínas que se podrían unir al ARN

Tabla 9. Dominios proteicos determinados por Pfam e Interpro en las 7 proteínas del secretoma experimental, con un PM < 40 KDa y de ubicación nuclear.

Uniprot ID	Nombre de la proteína	Tamaño en AA	Pfam		Interpro	
			Dominio	Inicio- fin	Dominio	Inicio- fin
P9WNC3	Formamidopirimidina-ADN glicosilasa 1	289	Fapy_DNA_glyco (Formamidopirimidina-ADN glicosilasa dominio N-terminal)	1-124	Dedo de zinc, tipo ADN glicosilasa / AP liasa	251-285
			H2TH (Dominio H2TH)	144-234	Dedo de zinc, tipo FPG / IleRS	257-286
			zf-FPG_IleRS (Dedo de zinc FPG IleRS)	257-286	Formamidopirimidina-ADN glicosilasa, dominio catalítico	1-125
					ADN glicosilasa / AP liasa, H2TH Unión al ADN	144-236
A0A089QKZ7	Proteína no caracterizada Rv1155A	62	DUF5302 (familia de función desconocida)	11-62	Proteína de función desconocida DUF5302	12-62

P9WH37	Proteína ribosómica 30S S3	274	KH_2 (dominio KH)	37-110	Dominio de homología K K Dominio de homología, tipo 2	59-124
			Ribosomal_S3_C (Proteína ribosómica S3, dominio C-terminal)	119-202		Proteína ribosómica S3, C-terminal
B5TV88	Proteína similar a ESAT-6	100	WXG100 (Proteínas de 100 residuos con WXG)	3-88	NO	
P9WKN9	Proteína no caracterizada Rv0942	92	NO		NO	
P9WFL1	UPF0232 proteína Rv0004	187	DciA (antecedente de ADN)	75-162	Proteína de función desconocida DUF721 / UPF0232 (familia)	3- 174
I6X642	Proteína no caracterizada	94	NO		NO	.

	Proteínas que se podrían unir al ADN
	Proteína que se podría unir al ARN y ADN

Para complementar la información de la ontología génica y los dominios proteicos, se mapearon regiones con afinidad por el ADN en las proteínas de interés, mediante el programa DNABIND. De las 36 proteínas del secretoma teórico, con ubicación nuclear y PM <40 KDa, un total de 21 (58.3%) tuvieron un resultado positivo, es decir, potencialmente podrían unirse al ADN (Tabla 10). Dentro de ellas estaban la 8-oxo- (d) GTP fosfatasa, la posible proteína similar a histona Hns, la proteína ribosómica 50S L33 2, metanotiol S-metiltransferasa, entre otros (Tabla 10).

Tabla 10: Probabilidad de unión al ADN de las proteínas del secretoma teórico, nucleares y con PM<40 KDa (n=36), usando el programa DNABIND.

Uniprot ID	Nombre de la proteína	Parámetros		
		Puntuación	Probabilidad de unión al ADN	Predicción
P9WH95	Proteína ribosómica 50S L33 2	8.722	0.9998	Unión al ADN
P95225	Antitoxina putativa Rv0268c	8.269	0.9997	Unión al ADN
P9WM27	Proteína no caracterizada Rv1322	7.204	0.9993	Unión al ADN
P9WLI5	Proteína no caracterizada Rv2206	6.436	0.9984	Unión al ADN

P9WL11	Proteína no caracterizada Rv3073c	5.161	0.9943	Unión al ADN
O53222	Toxina similar a HicA	4.304	0.9867	Unión al ADN
L0T8G6	Proteína no caracterizada	4.216	0.9855	Unión al ADN
O53621	Probable proteína de membrana	4.171	0.9848	Unión al ADN
I6X7F9	Regulador transcripcional Rv3488	3.926	0.9807	Unión al ADN
I6YHB0	Posible proteína similar a histona Hns	3.715	0.9762	Unión al ADN
L0T9U5	Probable proteína del fago PhiRv1	3.587	0.9731	Unión al ADN
O53402	Integrasa probable	3.233	0.962	Unión al ADN
P96376	Posible proteína de membrana conservada	3.16	0.9593	Unión al ADN
O53401	Proteína no caracterizada	2.867	0.9462	Unión al ADN
P95205	Proteína conservada de la familia de repeticiones 13E12	2.481	0.9228	Unión al ADN
O06213	Probable proteína de membrana rica en prolina conservada	2.349	0.9129	Unión al ADN
P9WLF7	Proteína no caracterizada Rv2271	2.133	0.8941	Unión al ADN
I6XWB9	Proteína que contiene el dominio DUF3073	1.887	0.8684	Unión al ADN
I6Y8X8	Transposasa putativa	1.723	0.8485	Unión al ADN
P9WIY3	8-oxo- (d) GTP fosfatasa	1.518	0.8203	Unión al ADN
O05883	Metanotiol S-metiltransferasa	0.685	0.6649	Unión al ADN
P9WIR7	Proteína secretada rica en alanina y prolina Apa	-2.288	0.0921	Sin unión al ADN
O05574	Proteína rica en serina conservada	-0.053	0.4867	Sin unión al ADN
P9WK79	Lipoproteína putativa LppB	-0.603	0.3537	Sin unión al ADN
P9WMT9	Factor de elongación de la transcripción GreA	-0.828	0.304	Sin unión al ADN
I6X562	Proteína que contiene el dominio DUF3039	-1.087	0.2521	Sin unión al ADN
P9WNK5	Proteína EsxB similar a ESAT-6	-1.262	0.2206	Sin unión al ADN
O53937	Ferredoxina probable	-2,436	0.0805	Sin unión al ADN
O06170	Posible proteína de membrana rica en prolina conservada	-2.505	0.0755	Sin unión al ADN
O53572	Proteína de membrana probablemente conservada	-2.724	0.0616	Sin unión al ADN
O06555	Proteína rica en ala conservada	-3.337	0.0343	Sin unión al ADN
I6YFL7	Rubredoxina	-3.988	0.0182	Sin unión al ADN
I6YGY9	Proteína rica en prolina hipotética conservada	-4.64	0.0096	Sin unión al ADN
P9WJC3	Proteína EspJ asociada a secreción de ESX-1	-4.792	0.0082	Sin unión al ADN
O05447	Proteína no caracterizada	-5,394	0.0045	Sin unión al ADN
P9WIR9	Homólogo de proteína antigénica de 34 kDa	-6.698	0.0012	Sin unión al ADN

Por su parte, 5 de las 7 proteínas del secretoma experimental (71,4%) tuvieron un resultado positivo, sugiriendo que podían asociarse al ADN (Tabla 11). Dentro de ellas, se encontraban la formamidopirimidina-ADN glicosilasa 1, la Proteína ribosómica 30S S3, la UPF0232 proteína Rv0004, entre otros (Tabla 11).

Tabla 11: Probabilidad de unión al ADN de las proteínas de secretoma experimental, nucleares y con PM<40 Kda (n=7), usando el programa DNABIND.

Uniprot ID	Nombre de la proteína	DNABIND		
		Puntuación	Probabilidad de unión al ADN	Predicción
A0A089QKZ7	Proteína no caracterizada Rv1155A	4.729	0.9912	Unión de ADN
P9WH37	Proteína ribosómica 30S S3	3.566	0.9725	Unión de ADN
P9WNC3	Formamidopirimidina-ADN glicosilasa 1	3.362	0.9665	Unión de ADN
P9WFL1	UPF0232 proteína Rv0004	3.236	0.9622	Unión de ADN
P9WKN9	Proteína no caracterizada Rv0942	2.222	0.9022	Unión de ADN
I6X642	Proteína no caracterizada	-0,195	0.4514	sin unión al ADN
B5TV88	Proteína similar a ESAT-6	-1.262	0.2206	sin unión al ADN

4.6. Estudio de homología con la genotoxina CdtB

Los resultados positivos de la etapa anterior mostraron que *M. tuberculosis* posee proteínas de secreción dirigidas al núcleo que podrían unirse a ácidos nucleicos. Considerando este hallazgo, nos propusimos averiguar si dichos factores son homólogos de la genotoxina CdtB, la cual daña el ADN. Mediante una prueba en BLAST se averiguó si alguna de las 43 proteínas del secretoma teórico y experimental de *M. tuberculosis*, dirigidas al núcleo y de PM<40 KDa, eran homólogas con la CdtB (subunidad B de la toxina de distensión citoletal) de *E. coli*, *S. typhi* y *C. hyointestinalis*. El resultado de BLAST fue negativo a homología indicando que, dentro del proteoma de *M. tuberculosis*, no existen homólogos de las CdtB de *E. coli*, *S. typhi* y *C. hyointestinalis*.

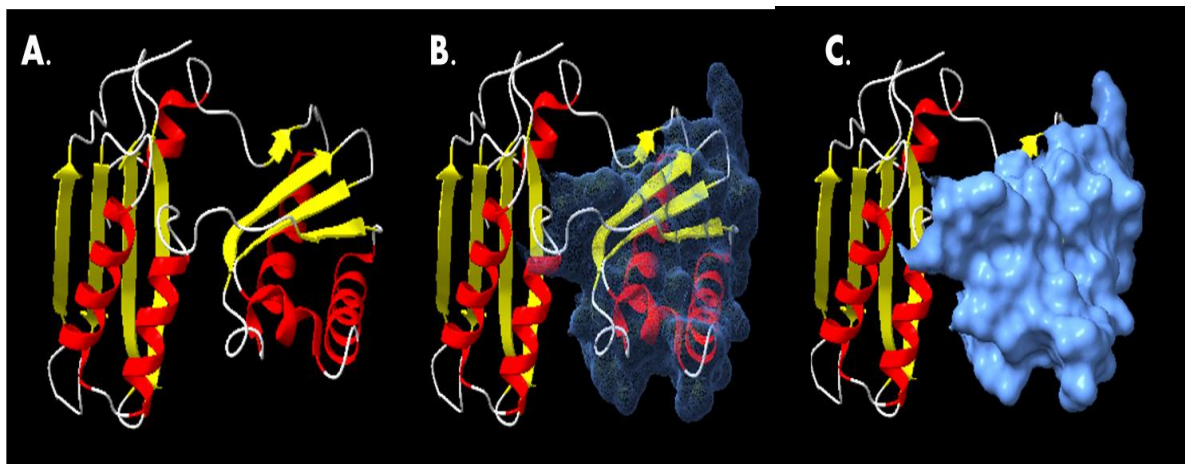
4.7. Modelamiento y visualización molecular

Para la etapa de modelamiento y simulaciones de docking proteína-ADN, se seleccionaron sólo aquellas proteínas cuyo resultado en los tres análisis realizados (ontología génica, exploración

de dominios proteicos y predicción de unión ácidos nucleicos) fueran positivos, es decir, que por su naturaleza poseían la capacidad de interactuar con el ADN/ARN. Siguiendo dicho criterio, se eligieron la proteína reguladora transcripcional Rv3488 (I6X7F9), la proteína S3 ribosomal 30S (P9WH37) y la Formamidopirimidina-ADN glicosilasa (P9WNC3).

A continuación, se realizó la visualización de dichas proteínas. Dos de ellas, la proteína reguladora transcripcional Rv3488 (I6X7F9) y la proteína S3 ribosomal 30S (P9WH37), contaban con su estructura experimental resuelta en el PDB (Protein Data Bank), mientras que la Formamidopirimidina-ADN glicosilasa (P9WNC3) no contaba con una estructura resuelta.

La estructura 3D de la proteína S3 ribosomal 30S de *M. tuberculosis* (código PDB: 7KGB) se encuentra formando parte de un complejo ribosomal, la cual fue determinada mediante microscopía electrónica, con una resolución de 2.7 Å (67). Se recuperó y descargó únicamente



la proteína de interés, la S3 ribosomal 30S, la cual es un monómero de 274 residuos (Fig. 6A). Su estructura está compuesta por 4 α -hélices bien constituidas y 2 pequeñas hélices, con presencia de 7 láminas β -plegadas (Fig. 6A). El dominio proteico KH2, comprendido entre los residuos 37 al 110 (Tabla 9), está involucrado en la unión a los ácidos nucleicos (ARN o ssADN) y está conformada por 3 láminas β -plegadas y por 3 α -hélices (Fig. 6B-C).

Figura 6. Estructura 3D de la proteína S3 ribosomal 30S (ID Uniprot: P9WH37) (PDB: 7KGB). **A:** Muestra la estructura secundaria de la proteína (α -hélices: rojo, láminas- β : amarillo). **B** y **C** muestran una representación semisólida y sólida respectivamente, del dominio KH (color celeste) que reconoce ácidos nucleicos.

La estructura de la proteína reguladora transcripcional Rv3488 (código PDB: 5ZHC), también está resuelta por difracción de rayos-X, con una resolución de 1.97 Å, como un homodímero de 226 aminoácidos (68) (Fig. 7A). Cada subunidad del dímero, está compuesta por 4 α -hélices y 2 láminas β plegadas (Fig. 7A-B) (68). En la estructura apreciamos también 6 aminoácidos adicionales en el N-terminal (-5 a -1), agregados sobrantes de la expresión y purificación de la proteína (68). El dominio PadR, presente en ambos monómeros, es responsable de la unión al ADN y se encuentra entre el residuo 13 al 83 (Tabla 8). Estructuralmente, el dominio PadR está comprendido por 3 α -hélices y 2 hojas β plegadas (Fig. 7C-D).

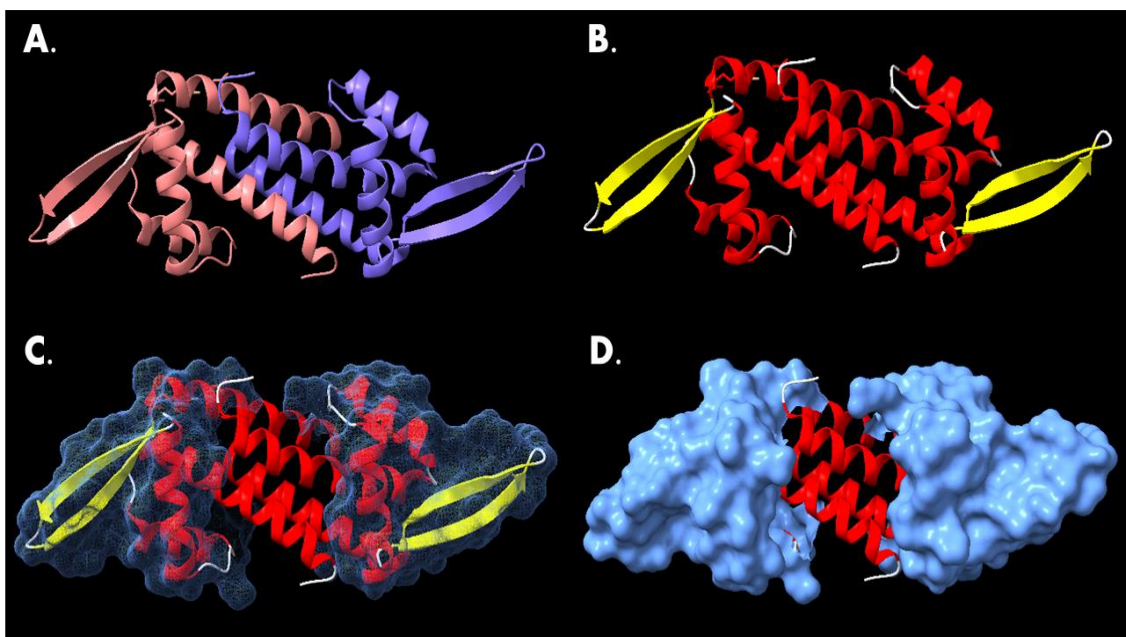


Figura 7. Estructura 3D de la proteína reguladora transcripcional Rv3488 (ID Uniprot: I6X7F9) (PDB: 5ZHC). **A:** estructura 3D del homodímero (los monómeros se han representado de color rosado y azul). **B:** elementos de estructuras secundarias presentes en el complejo proteico (α -hélices: rojo, β -láminas: amarillo, loops: blanco). **C** y **D** muestran una representación semisólida y sólida respectivamente, del dominio PadR, presente en cada subunidad (color cian).

Por otro lado, dado que la estructura de la Formamidopirimidina-ADN glicosilasa (P9WNC3) no estaba disponible en el PDB, el modelo 3D fue obtenido de AlphaFold (AF-P9WNC3-F1) (Fig. 8). En el visor 3D de AlphaFold, se observó un alto nivel de confianza del modelo en casi

toda la estructura, excepto en las regiones comprendidas entre Thr69 y Thr76 y entre Gly228 y Arg248, donde el nivel de confianza del modelamiento fue bajo o muy bajo (Fig. 8).

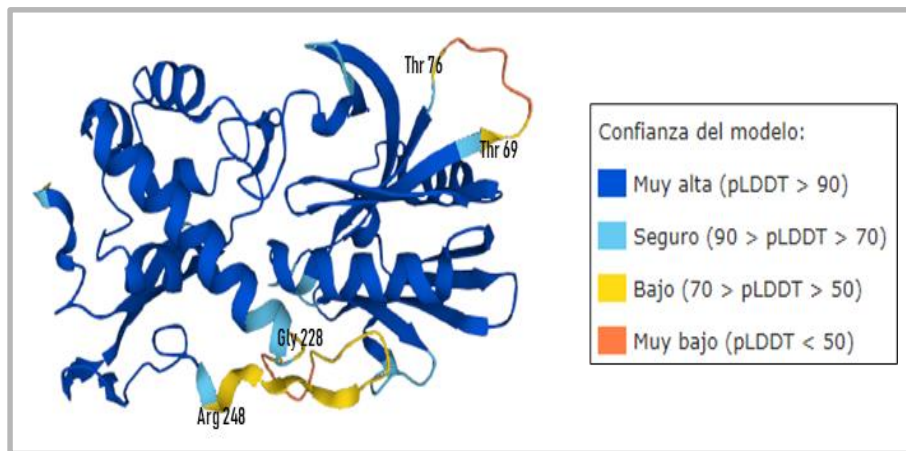


Figura 8. Estructura 3D de la proteína formamidopirimidina-ADN glicosilasa modelada por AlphaFold (ID AlphaFold: AF-P9WNC3-F1). La imagen muestra la calidad del modelo en base a la puntuación de confianza por residuo (pLDDT), de entre 0 y 100.

Las regiones con baja calidad no corresponden al dominio dedos de Zinc, pero sí a una porción al dominio H2TH (helix-2turn-helix domain) que afecta a los residuos de Gly228 y Tyr236. Para mejorar la calidad del modelo de AlphaFold, la estructura fue sometida a refinamiento energético con DeepRefiner, obteniéndose 5 modelos. El modelo elegido fue aquel con mayor puntuación de calidad global, según el cómputo de DeepRefiner, el cual correspondió al primer modelo de la lista resultante, el más favorecido energéticamente (Anexo 6, Fig. 9A). Al comparar la estructura obtenida de AlphaFold con la refinada, se obtuvo un RMSD de 0,51 Å entre los 289 residuos, que sugiere que la conformación se reajustó, tal como se esperaba después del refinamiento.

La estructura de la Formamidopirimidina-ADN glicosilasa está compuesta por 3 α -hélices bien constituidas y 3 pequeñas α -hélices, con presencia de 10 láminas β -plegadas largas y 4 pequeñas (Fig. 9A). El dominio de unión al ADN, H2TH, está comprendido entre el residuo 144 y 236 y formado por 2 α -hélices bien constituidas y 2 pequeñas hélices, mientras que el dominio dedos de Zinc comprende del residuo 257 al 286 (Tabla 9), y su estructura está conformada por 3 láminas β plegadas (Fig. 9B-C).

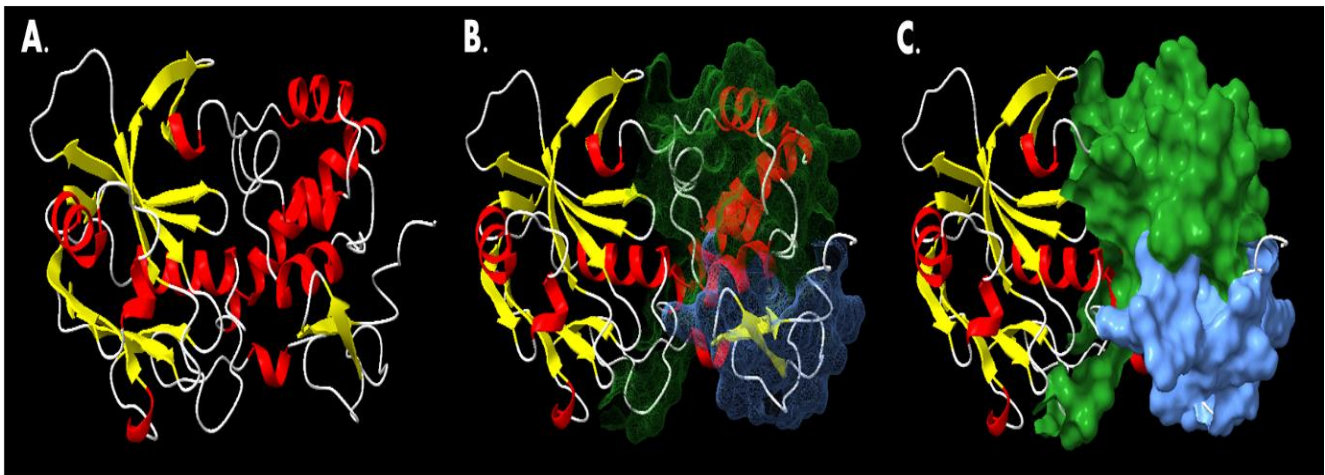


Figura 9. Estructura 3D refinada de la proteína formamidopirimidina-ADN glicosilasa (ID Uniprot: P9WNC3). **A:** muestra los elementos de estructura secundaria (α -hélices: rojo, β -laminas: amarillo, loops: blanco). **B y C** muestran una representación semisólida (B) y sólida (C) de los dominios que participan en la unión al ADN (dedos de zinc: celeste, dominio H2TH: verde).

Al comparar los dominios de unión al ácido nucleico entre las 3 proteínas analizadas, observamos que el tamaño de dichas regiones varía entre las tres proteínas de secreción analizadas. Así, la Formamidopirimidina-ADN glicosilasa es la que dispone de un área más extensa para la interacción con el ADN, incluyendo un dominio de 92 residuos (H2TH) y otro de 29 residuos (Dedos de Zinc). En tanto, la proteína reguladora transcripcional Rv3488 tiene un dominio (PadR) de 70 residuos, mientras que la proteína S3 ribosomal 30S (que une ARN) tiene un dominio (KH) de 73 residuos (Tablas 7 y 8).

4.8. Docking Molecular

Una vez descritas, a nivel estructural, las 3 proteínas de secreción de *M. tuberculosis*, que tendrían como destino el núcleo de la célula huésped, se simuló el acoplamiento de dichas proteínas con una secuencia de ADN/ARN, aplicando el algoritmo de HDock. Primero se realizó el *docking* de la proteína ribosómica 30S S3 (P9WH37, PDB: 7KGB) con una cadena simple de ARN (25 pb), el cual es su ligando natural. La secuencia de ARN fue extraída del complejo de la proteína Nova1 con ARN (5'-CCpCGCGGAUCAGUCACCCAAGCGAG-3'), disponible con código PDB 2ANR (69). La elección de esta secuencia de ARN para la simulación obedeció al hecho de que está resuelta en complejo con el dominio KH de Nova 1,

siendo KH también el dominio presente en la proteína ribosómica 30S S3 (69). El complejo resultante (Fig. 10 A-B) obtuvo una puntuación de acoplamiento o energía de estabilización de -306.39 kcal/mol (Anexo 7). Los resultados muestran que una parte de los residuos que participan en la unión al ARN están dentro del dominio KH (Arg48, Ala70, Arg 71, Pro 72, Gly73, Ile74, Glu104, Val105, Lys106, Asn107, Pro108 y Glu109), mientras que otros están fuera de dicho dominio (Gly118, Glu121, Gln122, Ser124, Asn125, Arg126, Val127, Lys135, Ala136, Ser139, Arg142 y Gln143) (Fig. 10 C-D). En cuanto a las interacciones intermoleculares entre la proteína y el ARN, de acuerdo al programa PLIP, se forman enlaces de hidrógeno, interacciones catiónicas- π y puentes salinos. (Anexo 8)

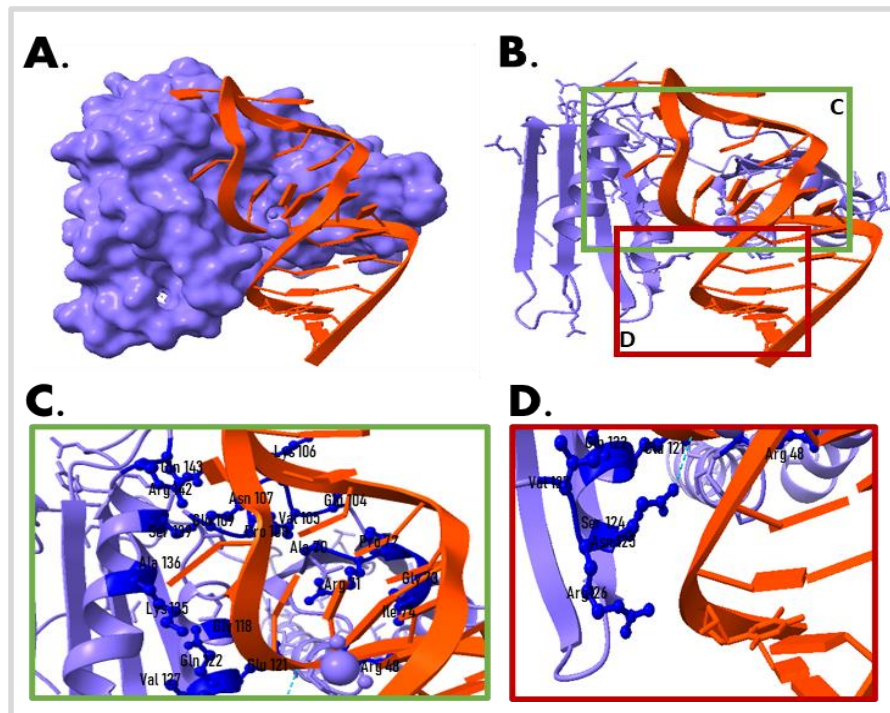


Figura 10. Complejo modelado entre la proteína ribosómica 30S S3 (P9WH37) con ARN. A y B muestran el mejor modelo de acoplamiento predicho por el servidor HDOCK donde se visualiza la subunidad proteica de azul y la cadena de ARN naranja. C y D muestran todos los residuos de la proteína que participan en la unión al ARN.

Luego se simuló la interacción de la proteína reguladora transcripcional Rv3488 (I6X7F9, PDB: 5ZHC) con una porción de ADN de doble hebra de 28 pb. La secuencia fue extraída del complejo entre el dominio PadR de *Bacillus subtilis* en complejo con ADN de doble hebra (5'-CGGAACATGTAAATAGTTACATGATTAC-3') disponible con el código PDB 5X11 (70). Nuevamente, la elección de esta secuencia se basó en el hecho de que el dominio PadR está

presente en la proteína reguladora transcripcional Rv3488 (70). Al finalizar la rutina, se eligió el complejo modelado proteína-ADN considerando aquel con mejor puntuación de acoplamiento (menor energía de afinidad), eligiendo aquel con un valor de -278.11 kcal/mol (Fig. 11 A-B, Anexo 7). La interacción con el ADN ocurre en el N- y C-terminal y en la interfaz de dimerización (Fig. 10 A-B). Observamos que algunos residuos que median la unión al ADN pertenecen al dominio PadR, en la subunidad A, incluyendo His23, Gly24, Ala25, Trp26, Gly41, Thr42, Tyr44, Pro45, His48, Arg49, Ser58, Gly65, Arg66, Arg68, Arg69, Val70 y Tyr71. Otros residuos de la subunidad B también intervienen en la unión al ADN, incluyendo Ser39, Thr42, Thr46, Arg49, Gln60, Arg66, Ala67, Arg68 y Arg69 (Fig. 11 C-E). Sin embargo también se identificó otros están fuera de dicho dominio (Arg2, Glu3, Arg6 de la subunidad A y Arg2, Glu3, Arg6 de la subunidad B). Al evaluar las interacciones intermoleculares entre la proteína y el ADN, apreciamos que predominan los puentes de hidrógeno, puentes salinos e interacciones hidrofóbicas (Anexo 9).

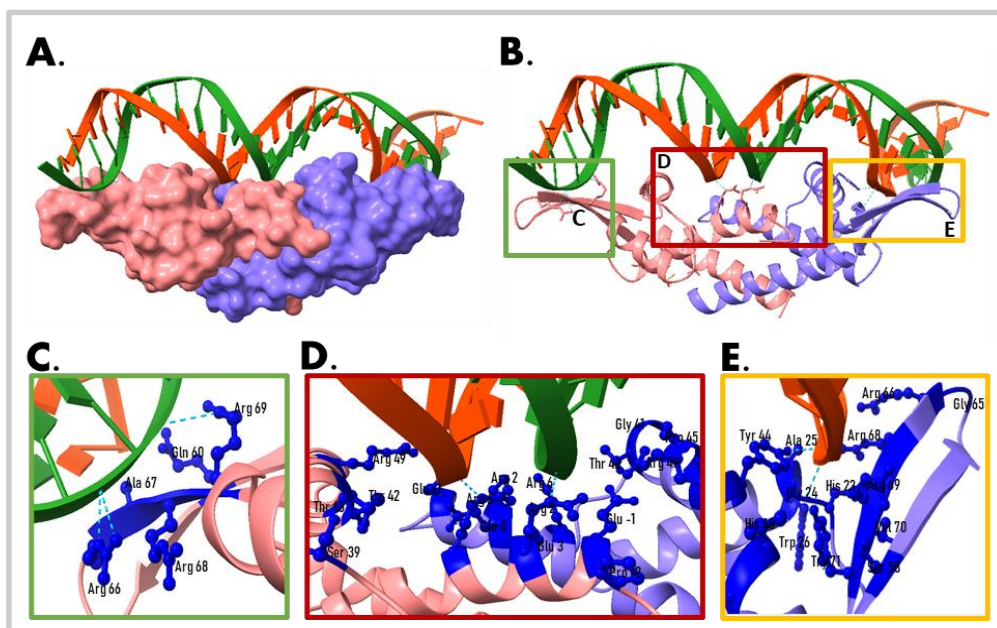


Figura 11. Complejo modelado de la proteína reguladora transcripcional Rv3488 (I6X7F9) con un ADN de doble hebra. A y B muestra el mejor modelo de acoplamiento predicho por el servidor HDock donde se visualiza la subunidad A (azul) y la subunidad B (rosado), cadenas de ADN en naranja y verde. **C, D y E:** muestran todos los residuos de la proteína involucrados en la unión al ADN. En C, las líneas entrecortadas representan puentes de hidrógeno intermoleculares.

Finalmente, el acoplamiento de la formamidopirimidina-ADN glicosilasa de *M. tuberculosis* (P9WNC3) con ADN se simuló usando una cadena de ADN de doble hebra de 13 pb (5'-GGCTTCCTCCTGG-3') extraída del complejo de la homóloga en *E. coli* con ADN (PDB: 1K82) (71). De los complejos modelados proporcionados por HDOCK, se seleccionó el modelo 1 (Fig. 12 A-B), el cual fue el mejor en términos de puntuación de acoplamiento, con un valor de -571.27 kcal/mol (Anexo 7). Una parte de los residuos que participan en la unión al ADN pertenecen a los dominios de dedos de Zinc y H2TH (Fig. 12 C-D), incluyendo a Lys165, Ser167, Arg171, Gln172, Asp 175, Gln176, Ile182, Gly183, Asn184, Ile185, Ser234, Leu235, Tyr236, Arg271, Phe272, Met273, Asn274, Arg275 y Ser276. Otros también interactúan con el ADN, pero están fuera de dichos dominios, incluyendo Pro2, Glu3, Arg32, Ala33, Val34, Arg35, Lys61, His83, Gly85, Met86, His100, Gln118, Arg119, Thr120, Phe121, Ser135, Val137, Pro138, Pro140, Val237, Asn238, Val239, Asn240 y Tyr253 (Fig. 12 C-D). De acuerdo a los resultados de PLIP, las interacciones que permiten la unión al ADN son nuevamente puentes de hidrógeno, puentes salinos, interacciones hidrofóbicas y una interacción catiónica π (Anexo 10).

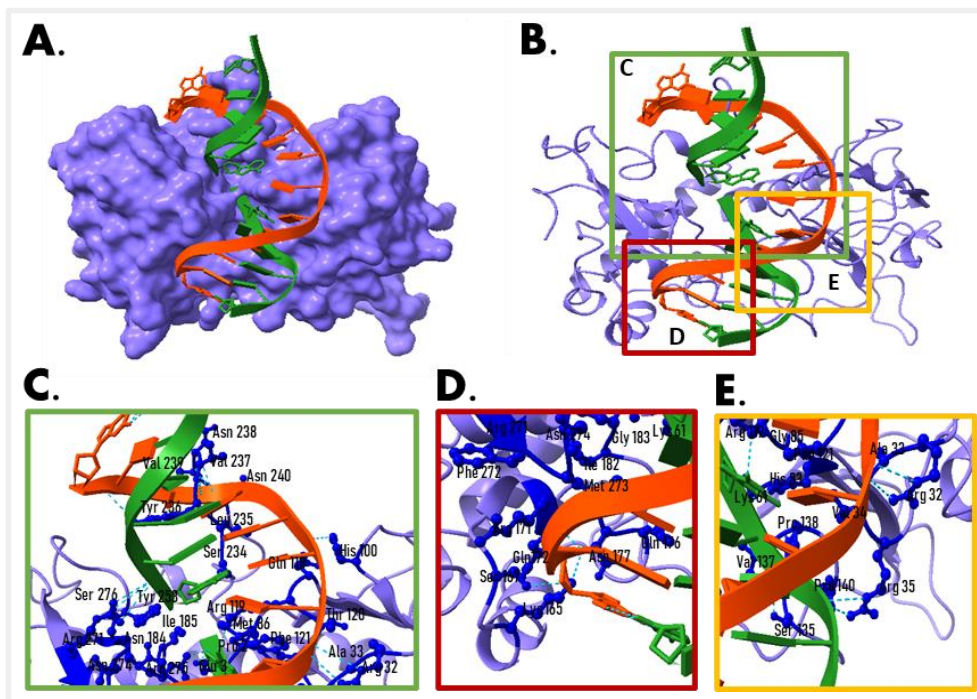


Figura 12. Modelo del complejo de la proteína Formamidopirimidina-ADN glicosilasa (P9WNC3) con ADN de doble hebra. A y B muestra el mejor modelo de acoplamiento predicho por el servidor HDOCK donde se visualiza la proteína en azul y las cadenas de ADN en naranja y verde. C, D y E: muestran todos los residuos de la proteína que participan en la unión al ADN.

5. DISCUSIÓN

Existen indicios de que las personas que padecen tuberculosis presentan un alto riesgo de desarrollar cáncer de pulmón (CP) (18-22). Sin embargo, se desconoce si *M. tuberculosis* estaría propiciando el desarrollo de CP, y en caso así fuera, tampoco se tiene noción sobre el mecanismo que podría estar implicado en ello (1,3,7). Por ello, en el presente trabajo, buscamos estudiar si *M. tuberculosis H37Rv* podría presentar proteínas de secreción que se dirijan al núcleo de la célula huésped y que tengan la capacidad de interactuar con el ADN, pudiendo ser éste un potencial mecanismo de daño al material genético. Nos valimos del proteoma y secretoma de *M. tuberculosis H37Rv*, determinados experimentalmente, para predecir proteínas de secreción que se dirijan al núcleo, considerando sus propiedades fisicoquímicas y moleculares. Se encontraron 58 proteínas de secreción destinadas al núcleo de la célula huésped, de las cuales 3 mostraron tener una fuerte interacción con ácidos nucleicos, al simular su acoplamiento con ARN/ARN.

La búsqueda de proteínas de secreción dirigidas al núcleo no es una actividad trivial. De hecho, algunas bacterias patógenas como *H. pylori* y *Listeria monocytogenes*, secretan factores de virulencia dirigidos al núcleo de la célula infectada, las cuales pueden afectar la integridad del genoma, siendo un ejemplo las genotoxinas (1,7-11, 72,73). En este sentido, el internamiento de proteínas bacterianas al núcleo de las células eucariotas podría causar daño al material genético lo cual, a la larga, puede desencadenar la transformación (74). Este mecanismo ha sido de interés en diversos estudios *in silico*, motivados por averiguar si la carcinogénesis asociada con infecciones bacterianas puede tener su explicación en la presencia de proteínas de secreción transportadas al núcleo del huésped (34-39). En dichas investigaciones se predijeron algunos factores de secreción bacterianos, con potencial destino nuclear, incluidos la ribonucleasa R y el regulador transcripcional BtsR de *S. typhimurium*, la chaperona DnaJ de *C. pneumoniae*, la topoisomerasa 1 de ADN de *H. pylori* y la proteína de partición cromosómica Smc de *M. hominis* (34-39). La existencia de dichos polipéptidos en *M. tuberculosis* es aún desconocida y fue motivo de investigación en la presente tesis.

La identificación de factores de secreción con destino nuclear, predichos en el presente trabajo, fue posible gracias a que se contaba con el proteoma y secretoma de *M. tuberculosis* determinados experimentalmente (27-33). Sin embargo, de las 837 reportadas en el secretoma, solo 133 lograron ser recuperadas, resultando que el resto no pertenecían al proteoma de *M. tuberculosis H37Rv* sino a otras micobacterias e incluso a otros microorganismos. Asimismo,

sucedió el caso en el que el identificador Uniprot o del GenBank no pudo ser reconocido. No encontramos una explicación a esta observación, pero es claro que más de 700 proteínas del secretoma experimental no pertenecen a la cepa *H37Rv* como se informó en el artículo e incluso no se encuentran disponibles en las bases de datos consultados (33). Sin embargo, podría deberse a que esta base de datos se encuentra en constante actualización y reajuste de su base de datos e incluso pudo haberse realizado una limpieza o curación por lo que los números de GI y Uniprot ID de dichas proteínas fueron removidos.

Dado que no fue posible el análisis del secretoma experimental completo, decidimos correr nuestro propio algoritmo para la predicción teórica del secretoma de *M. tuberculosis* a partir del proteoma integrado, conformado por resultados de 6 estudios (27-32). Como primer hallazgo, se predijo que 972 proteínas son secretadas por la micobacteria, de las cuales, 23 son comunes con el secretoma experimental recuperado. La secreción de proteínas al medio extracelular es un proceso ligado a la patogénesis y virulencia de la micobacteria, que ocurre a través de distintas vías, dependiendo de la presencia de un péptido señal (1,7).

En el presente estudio, se observó que el 35% (n=342) del secretoma predicho de *M. tuberculosis* posee un péptido señal. Ello se determinó usando un algoritmo previamente empleado por Tucci P. et al (27), para predecir 140 proteínas de secreción en *M. tuberculosis*. De hecho, dichos factores previamente identificados, estuvieron incluidos en el grupo de 342 polipéptidos de *M. tuberculosis* determinados en nuestro trabajo (con péptido señal). Precisamente, esas proteínas serían secretadas por las vías clásicas de secreción al contar con un péptido señal. Acorde a la literatura, corresponderían a la vía de secreción general conservada Sec (SecA1) y las vías de translocación de arginina gemela (Tat) reportadas en *M. tuberculosis* (75, 76). Siendo estas vías las que realizan la mayor parte de la exportación de las proteínas de la micobacteria (76). Los resultados también mostraron 630 proteínas sin péptido señal, que serían secretadas por *M. tuberculosis* a través de las vías no clásicas de secreción o exportadores especializadas, como el sistema de exportación ESX y SecA2. Si bien este último sistema exporta proteínas carentes de péptido señal, también parece estar implicado en la secreción dependiente de un péptido señal (76,77). Más aún, SecA2 mediaría la exportación de proteínas que dañan el material genético (24), sin embargo, en el presente trabajo no se ha podido determinar mediante los programas usados qué proteínas serían secretadas por SeacA2.

Una vez identificados los polipéptidos que componen el secretoma experimental y el predicho, se determinó su localización subcelular. Se emplearon diferentes herramientas bioinformáticas

basadas en máquinas de vectores de soporte (SVM), que han sido previamente utilizadas y validadas en otros estudios *in silico* en proteínas bacterianas, por lo que resultan convenientes para los fines de esta investigación (34-39, 46-49). La utilización de hasta 4 programas en esta etapa otorga robustez al algoritmo de predicción de localización subcelular. Ello sumado al hecho de imponer el criterio de destino nuclear computado por al menos dos predictores, hacen de este un método confiable y preciso. Se determinó que dentro del secretoma de *M. tuberculosis* existen un total de 58 proteínas destinadas al núcleo de la célula huésped. Hasta donde sabemos, esta es la primera vez que se determina que una porción del secretoma de *M. tuberculosis* está destinado al núcleo de la célula eucariota. Exactamente el 7.5 % concierne al secretoma experimental y el 4.93 % al secretoma teórico determinado en la presente tesis.

Para mejorar la precisión de estas determinaciones teóricas, a continuación, se determinó el PM y PI de las 58 proteínas reconocidas en la etapa anterior. En cuanto al PI, se observó un patrón de distribución similar al reportado en proteínas secretoras predichas nucleares en *Salmonella typhimurium*, *C. pneumoniae*, *H. pylori*, *Mycoplasma hominis* y *E. coli* (34, 35, 37-39). En *M. tuberculosis* destacaron un grupo de 26 proteínas con altos valores de PI (incluso mayores a 8), siendo polipéptidos potencialmente dirigidos al núcleo y/o nucleolo. Esta afirmación se apoya en el hecho de que las proteínas con carga positiva y con altos valores de PI, tienden a encontrarse y/o acumularse en mayor cantidad en el núcleo de las células (66,78). Además del PI, nuestro estudio puso especial énfasis en el PM ya que uno de los mecanismos de ingreso al núcleo depende del tamaño de la proteína. De este modo, las proteínas con PM <40 KDa son capaces de internarse en el núcleo por medio de la difusión pasiva (52). Mientras que las proteínas con un PM > 40 KDa tienden a ingresar por medio de los poros nucleares, dependiendo de una señal de localización nuclear (NLS), tal como se investigó en la bacteria *H. pylori* (79). En la presente tesis, se seleccionó únicamente las proteínas con un PM < 40 KDa, es decir, aquellas que pueden difundir pasivamente al ambiente nuclear (n=43). La micobacteria podría tener otras proteínas de secreción dirigidas al núcleo vía NLS, pero las mismas no fueron investigadas en este trabajo. Sin embargo, se recomienda considerarlos en futuras investigaciones.

La secreción de proteínas por parte de bacterias, cuyo destino es el núcleo, típicamente se relaciona con genotoxicidad, es decir, daño en el material genético. Diversas genotoxinas han sido descritas y vinculadas a la patogenicidad bacteriana, como la CDT, la colibactina y la Usp (9-11). Sin embargo, hasta donde sabemos, no se han descrito genotoxinas en *M. tuberculosis*,

aunque sí se conoce el daño en el ADN que causa dicha bacteria en los macrófagos humanos y similares efectos genotóxicos en *M. leprae* (23, 24, 80). Es razonable pensar que los efectos genotóxicos de la micobacteria, en la célula eucariota, pueden estar mediados por factores proteicos tal como sucede en *E. coli*, *Klebsiella pneumoniae* y otras enterobacterias, por lo que nos propusimos identificar genotoxinas en *M. tuberculosis* a través del reconocimiento de homologías por BLAST. Dado que no existen secuencias consenso de las genotoxinas bacterianas, se realizó un análisis de homología únicamente con la Cdtb, que es la más común y conocida. El resultado negativo del BLAST sugiere que ninguna de las 43 proteínas de secreción dirigidas al núcleo (con PM < 40KDa) de *M. tuberculosis* presenta homología con Cdtb de *E. coli*, *S. typhi* ni *C. hyointestinalis*. Si existe homología de esas 43 proteínas de *M. tuberculosis* con otras genotoxinas es algo que no se determinó en esta tesis, pero podría investigarse en un futuro para consolidar los hallazgos reportados.

A continuación, se determinó la ontología génica de las 43 proteínas secretoras dirigidas al núcleo, con el fin de determinar los procesos y funciones en los que están involucradas. Los resultados sugieren que estos factores secretados y potencialmente internalizados al núcleo eucariota regulan procesos claves como la traducción, la transcripción, la reparación del ADN, entre otros. Su participación en dichos procesos, dentro del núcleo eucariota, podría provocar la alteración o disrupción de la maquinaria molecular eucariota. De hecho, 8 de las 43 proteínas de interés, tienen dominios de unión a ácidos nucleicos por lo que, al internalizarse al núcleo de la célula huésped, pueden formar complejo con el material genético o bien con ARNs. Así mismo, su capacidad de unión al ADN fue predicha por el programa DNABIND, sugiriendo que un total de 26 podrían potencialmente reconocer y unirse al ADN. En algunas de ellas se ha descrito su rol en la fisiopatología de la tuberculosis, pero no su capacidad de unirse a ácidos nucleicos. Por ejemplo, el Factor de elongación de la transcripción GreA es importante para la supervivencia y resistencia a rifampicina y vancomicina en *M. tuberculosis* (81). Dentro de este grupo también se encuentra la 8-oxo-(d) GTP fosfatasa, cuya función en *M. tuberculosis* sería el crecimiento y la supervivencia en condiciones de estrés oxidativo (82). Esta última resulta un factor de interés pues al secretarse e internalizarse al núcleo, podría promover el crecimiento celular, el cual es un *hallmark* del cáncer. La acción conjunta de estas proteínas secretoras dirigidas al núcleo podría perturbar el programa natural de las células huésped, sea o no a través de su asociación con el ADN.

En particular, en esta tesis, investigamos la capacidad de algunas de las proteínas secretoras

destinadas al núcleo de unirse a ácidos nucleicos. Seleccionamos 3 proteínas de *M. tuberculosis* que resultaron tener la propiedad de unirse al ADN según su ontología, así como contar con dominios de unión al ADN y mostrar una alta probabilidad de unirse al ADN, predicho con DNABIND. Uno de esos factores fue la proteína S3 ribosomal 30S (rpsC), la cual dentro de la micobacteria es un constituyente estructural de la subunidad 30S del ribosoma (P9WH37) (83). Algo interesante con esta proteína es que no sólo fue predicha de secreción por el algoritmo de este trabajo, sino que fue también determinada experimentalmente de secreción en otro trabajo (33). Bajo ese contexto y considerando su naturaleza, es posible que se acompleje con los componentes de los ribosomas humanos. Sin embargo, dado que su destino sería el núcleo, dicho ensamblaje entre la rpsC y los ribosomas no ocurrirían necesariamente en el citoplasma, pudiendo ocurrir en el nucleolo, lugar de ensamblaje de los ribosomas. Esto sumado a que esta proteína tiene un PI alto de 10.1, característica que la hace tener mayor probabilidad de destinarse al núcleo y acumularse en el nucleolo (66,78)

La ontología de la rpsC sugiere que dicho factor participa en la traducción y que se asocia al ARNm y ARNr (83). Esto concuerda con evidencia previa que demuestra que la rpsC se une al ARNm en la subunidad 70 S del ribosoma y de este modo lo posiciona para la traducción; asimismo, esta proteína puede interactuar con el ARNr formando una red tridimensional (83). Ello estaría dándose gracias a la presencia del dominio proteico ribosomal S3 C-terminal y del dominio de homología K tipo 2 (KH2), el cual está presente principalmente en proteínas procariotas que regulan la transcripción y traducción, tal como en la rpsC (84).

Se sabe que el dominio KH2 es responsable de la unión de la rpsC con el ARN, el mismo que también podría mediar la interacción de la rpsC con una cadena monocatenaria de ADN (ADNmc). Sin embargo, no se ha determinado experimentalmente si dicha interacción con ADNmc es factible (84). En nuestra tesis, la rpsC fue simulada en interacción con una hebra de ARN y no con ADNmc como con las otras proteínas de estudio, ya que el ARN es su ligando natural (65). Se comprobó que el dominio KH2 está involucrado en la interacción con el ARN, participando un total de 24 residuos de la rpsC, la mitad de ellos ubicados en el dominio KH2. Nuestros resultados también concuerdan con la conformación del complejo rpsC-ARN descrito en la literatura, observándose que los ácidos nucleicos se unen a una de las caras del dominio KH2, por medio de las hélices α (84,85).

Si bien la rpsC no parece ser un agente genotóxico, ya que su ligando natural es ARN, su participación en la etiología del cáncer no se descarta. La rpsC podría unirse a algunos

componentes del ribosoma del huésped, como el ARN, alterando así el proceso de traducción normal (86). Ya que las características de la rpsC sugieren que su destino es el núcleo, hipotetizamos que su unión al componente ribosómico podría ocurrir en el nucleolo, lugar de ensamblaje de ribosomas. Otra posibilidad es que la rpsC no se acompleje con el ARN en el ribosoma, sino que se asocie con transcritos a nivel del núcleo. De hecho, se sabe que la rpsC eucariota (rpS3e) puede encontrarse en un estado libre (no asociada con ribosoma) y regular diversos procesos biológicos extraribosómicos, tal como la reparación del ADN y apoptosis (87). Un dato interesante es que las alteraciones en la expresión o mutaciones en proteínas de unión al ARN (RBPs) son causantes de cáncer (88). Por lo tanto, la presencia de una RBP de *M. tuberculosis* en el núcleo de la célula humana podría disrumpir los procesos biológicos normales e incentivar vías oncogénicas que promuevan la tumorigénesis y la transformación celular (88).

Por su parte, el regulador transcripcional Rv3488 (I6X7F9) también sería una proteína secretada por *M. tuberculosis* (sin evidencia experimental), con características compatibles de una proteína destinada al núcleo en la célula huésped. Puesto que además de tener PM<40 KDa, curiosamente posee un PI alto de 9.6 lo que apoya aún más su direccionamiento nuclear (66,78). La ontología génica de Rv3488 indica que su función molecular consiste en la unión a iones metálicos, lo que explica su participación en la desintoxicación de metales tóxicos en patógenos, tales como el Cadmio y Zinc (68). Además de metales, esta proteína puede unirse al ADN, al ser un regulador transcripcional. De hecho, Kumari M. et al (68) comprobó que Rv3488 se une a su propio promotor de 30 pb, lo cual sugiere una autorregulación de su expresión génica, tal como se ha descrito en otras bacterias como *Clostridium difficile* (89). Además, se propone que es plausible su unión a otras regiones promotoras como la del transportador Rv1999c (68). Su conformación activa es un homodímero, observándose que cada monómero contiene el dominio PadR, responsable de la unión al ADN (68, 90). Debido a lo mencionado, se simuló el acoplamiento molecular entre el homodímero del regulador transcripcional Rv3488 con el ADN. Nuestros resultados muestran, al igual que la estructura cristalizada (PDB: 5X11), que el ADN se asentó sobre las caras correspondientes al dominio PadR, atravesando diagonalmente a ambos monómeros e interactuando con el 85% de los residuos del dominio PadR de cada dímero (70).

Según los resultados de este trabajo, el Rv3488 sería secretado por la micobacteria y podría dirigirse al núcleo de la célula huésped (como los macrófagos, células dendríticas, neutrófilos,

células no mielocíticas, células epiteliales alveolares de tipo II), donde es factible su interacción con el ADN genómico (17). La formación de un complejo entre una proteína bacteriana y ADN genómico humano podría tener un efecto dual. Por un lado, podría ser el preámbulo de una lesión en el genoma, tal como sucede con la genotoxina CDT o la proteína específica uropatógena (Usp), que tienen actividad DNAsa (8,11). Particularmente, en Rv3488, su capacidad de dañar directamente al ADN promoviendo así la oncogénesis, no ha sido descrita. Por otro lado, su interacción con el ADN podría impedir la unión de los reguladores de la transcripción propias del huésped, tal como lo sugieren estudios *in silico* (34-37). Esto podría afectar la transcripción y expresión de ciertos genes, tal como se ha descrito en algunos reguladores transcripcionales vinculados a la etiología del cáncer, cuyas mutaciones incluyendo amplificaciones y sobreexpresión activaron ciertos oncogenes y silenciando algunos genes supresores de tumores (90,91).

La última proteína que se identificó y seleccionó fue la formamidopirimidina-ADN glicosilasa 1 (P9WNC3), la cual ha sido determinada de forma experimental por Putim C. et al (33) como un factor de secreción y predicha en esta tesis como una proteína dirigida al núcleo de la célula huésped. Esta predicción en base a su peso molecular, es apoyada por un PI alto de 9.9, por lo que su direccionamiento nuclear es posible (66,78). Según los resultados de su ontología, esta proteína presenta funciones que implican su interacción con el ADN, incluida la reparación del material genético. Esto último ocurre mediante la escisión de las bases dañadas por oxidación o agentes mutagénicos, lo cual ocurre en el dominio catalítico formamidopirimidina-ADN glicosilasa (92). En esta proteína también identificamos la presencia de los dominios H2TH y dedos de Zinc tipo FPG/ IleRS, los cuales son dominios comunes descritos en las ADN glicosilasas/AP (71, 93). Los dedos de Zinc son dominios de unión al ADN bien conocidos, y el dominio H2TH participa principalmente en el reconocimiento de secuencias del ADN (93). La interacción del ADN y la formamidopirimidina-ADN glicosilasa 1 también fue visualizada. Nuestros resultados muestran a la cadena de doble hebra acoplada en una hendidura en el centro de la proteína, igual a la posición de la proteína-ADN cristalizada de *E. coli* (71), y teniendo contacto con pocos residuos ubicados en la cara del dominio H2TH y dedos de Zinc.

Si bien la formamidopirimidina-ADN glicosilasa 1 contribuiría a la supervivencia de *M. tuberculosis*, al reparar su propio ADN dañado a causa de ROS y especies reactivas de nitrógeno (RNS) generados por los macrófagos (94), la glicosilasa podría desempeñar un rol opuesto cuando es secretada y transportada al núcleo de la célula huésped. Debido a que esta

proteína escinde las bases del ADN e introduce *nicks* en la doble hebra, podría causar una aberración en el proceso de reparación del ADN humano, dañando así el material genético, incluso pudiendo llevar a la ruptura de la doble hebra del ADN (71, 93-95). Este evento ha sido también hipotetizado en las proteínas de reparación como la topoisomerasa de *M. hominis*, así como en MutS y MutL de *C. pneumoniae* y en MutS de *S. typhimurium* (34,35,38). De este modo, la formamidopirimidina-ADN glicosilasa 1 de *M. tuberculosis*, mostraría indicios de poseer una capacidad genotóxica, por lo que podría participar en la etiología del CP.

Hasta donde sabemos, es la primera vez que se describen proteínas secretadas por *M. tuberculosis* que se dirijan al núcleo de las células humanas y su potencial capacidad de unir ácidos nucleicos. Considerando que *M. tuberculosis* es causante de una de las enfermedades infecciosas más prevalentes del mundo, es importante conocer el daño que la micobacteria puede causar a la larga en el organismo humano, por ejemplo, a través de la secreción de factores destinados al núcleo. Si dichas proteínas median la etiología del cáncer de pulmón es algo que deberá investigarse más profundamente.

6. CONCLUSIONES

- El análisis del proteoma integrado (n=3753) de *M. tuberculosis H37Rv* arrojó 972 proteínas de secreción, de las cuales 25 coincidían con el secretoma experimental. Asimismo, se identificaron 58 proteínas destinadas al núcleo de las células del huésped.
- Tres proteínas (la Proteína ribosómica 30S S3, la proteína reguladora transcripcional Rv3488 y la Formamidopirimidina-ADN glicosilasa) presentan afinidad por los ácidos nucleicos y cumplen funciones de formación de ribosomas, factor de transcripción y regulador de la reparación del ADN, respectivamente. Su ingreso al núcleo, predicho en este estudio, supondría la interferencia en procesos claves de la célula huésped, como la traducción, regulación de la expresión génica y reparación del ADN.
- Los resultados obtenidos en el enfoque *in silico* dado en este trabajo, para el estudio del presunto papel que *M. tuberculosis* tendría en la etiología del cáncer de pulmón, puede abrir nuevas vías de investigación, en torno al entendimiento que se tiene sobre este tema.

7. LIMITACIONES Y RECOMENDACIONES

- De las 837 proteínas del secretoma experimental, sólo se logró recuperar la secuencia de 133 polipéptidos, pese a consultar varias bases de datos. La mayoría de esas

proteínas resultaban vinculadas a otros microorganismos o bien sus números de GI y Uniprot ID no fueron reconocidos, por lo que no se pudieron tomar en cuenta en este estudio. No hallamos una explicación clara de dicha problemática.

- La identificación de homología se valió únicamente de la secuencia de la genotoxina Cdtb y no de varias secuencias consenso, ya que las mismas no existen ni en la literatura ni en repositorios web.
- En este estudio seleccionamos únicamente proteínas con PM < 40 KDa como las candidatas a internarse en el núcleo de la célula huésped. Sin embargo, el grupo de PM > 40 KDa podría contener también proteínas destinadas al núcleo, a través de otros mecanismos diferentes a la difusión pasiva, por lo que su estudio es factible en futuras investigaciones.
- Dada la naturaleza teórica de este estudio, los hallazgos requieren una validación experimental que confirme la internalización nuclear de las proteínas secretoras predichas en *M. tuberculosis* y su asociación con ácidos nucleicos.

REFERENCIAS BIBLIOGRÁFICAS

1. Aurélie Gagnaire, Bertrand Nadel, Didier Raoult, Jacques Neefjes and Jean-Pierre Gorvel. Collateral damage: insights into bacterial mechanisms that predispose host cells to cancer. *Nature Review Microbiology*. [Internet]. 2017. [Cited 2021 Apr 8]; 15(2): 109-128. Available from: <https://www.nature.com/articles/nrmicro.2016.171>
2. Ameri, Aliakbar. Environmental Causes of Cancer. *American International Journal of Cancer Studies* [Internet] 2019 [Cited 2021 May 3]. 1. 27-34. Available from: <https://acsjournals.onlinelibrary.wiley.com/doi/full/10.3322/caac.21660>
3. Cai Q, Yuan Z. Overview of Infectious Causes of Human Cancers. *Adv Exp Med Biol*. [Internet]. 2017. [Cited 2021 Apr 7]; 1018:1-9. Available from: <https://pubmed.ncbi.nlm.nih.gov/29052128/>
4. Van ElslandD, Neefjes J. Bacterial infections and cancer. *EMBO Rep*. [Internet]. 2018 Nov. [Cited 2021 Apr 25];19 (11):e46632. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6216254/>
5. Mentis AA, Boziki M, Grigoriadis N, Papavassiliou AG. Helicobacter pylori infection and gastric cancer biology: tempering a double-edged sword. *Cell Mol Life Sci*. [Internet].2019, Jul. [Cited 2021 Apr 10];76(13):2477- 2486. Available from: <https://pubmed.ncbi.nlm.nih.gov/30783683/>
6. Eyvazi S, Vostakolaei MA, Dilmaghani A, Borumandi O, Hejazi MS, Kahroba H, Tarhriz V. The oncogenic roles of bacterial infections in development of cancer. *Microb Pathog*. [Internet]. 2020, Apr. [Cited 2021 Apr 10];141: 104019. Available from: <https://pubmed.ncbi.nlm.nih.gov/32006638/>
7. Sheweita SA, Alsamghan AS. Molecular Mechanisms Contributing Bacterial Infections to the Incidence of Various Types of Cancer. *Mediators of Inflammation* [Internet]. 2020 Jul 8 [cited 2021 Sep 27];2020:1–10. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7366195/>
8. Rosadi F, Fiorentini C, Fabbri A. Bacterial protein toxins in human cancers. *Pathog Dis*. [Internet]. 2016, Feb. [Cited 2021 May 11]; 74(1): ftv105. Available from: <https://academic.oup.com/femspd/article/74/1/ftv105/2467585>
9. Rihtar E, Žgur Bertok D, Podlesek Z. The Uropathogenic Specific Protein Gene *usp* from *Escherichia coli* and *Salmonella bongori* is a Novel Member of the TyrR and H-NS Regulons. *Microorganisms*. [Internet]. 2020, Feb. [Cited 2021 May 16]; 26;8(3):330. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7142922/>

10. Pons BJ, Vignard J, Mirey G. Cytolethal Distending Toxin Subunit B: A Review of Structure- Function Relationship. *Toxins (Basel)*. [Internet]. 2019, Oct. [Cited 2021 May 14]; 12;11(10):595. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6832162/>
11. Wernke KM, Xue M, Tirla A, Kim CS, Crawford JM, Herzon SB. Structure and bioactivity of colibactin. *Bioorg Med Chem Lett*. [Internet]. 2020, Aug. [Cited 2021 May 18]. 1;30(15):127280. Available from: <https://pubmed.ncbi.nlm.nih.gov/32527463/>
12. Chang Y, Moore PS, Weiss RA. Human oncogenic viruses: nature and discovery. *Philos Trans R Soc Lond B Biol Sci*. [Internet]. 2017, Oct. [Cited 2021 May 18]. 19; 372(1732): 20160264. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5597731/>
13. Van Tong H, Brindley PJ, Meyer CG, Velavan TP. Parasite Infection, Carcinogenesis and Human Malignancy. *EBioMedicine*. [Internet]. 2017, Feb. [Cited 2021 May 19];15:12-23. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5233816/>
14. Claudia Machicado, Maria Pia Soto, Luis Felipe La Chira, Joel Torres, Carlos Mendoza, Luis A. Marcos. In Silico Prediction of Secretory Proteins of *Opisthorchis viverrini*, *Clonorchis sinensis* and *Fasciola hepatica* that Target the Host Cell Nucleus. *Heliyon*. [Internet]. 2021. [Cited 2021 Jun 21]; 2405-8440. Available from: <https://www.sciencedirect.com/science/article/pii/S2405844021013074>
15. Furin J, Cox H, Pai M. Tuberculosis. *Lancet*. [Internet]. 2019 Apr. [Cited 2021 May 21]. 20;393(10181):1642-1656. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/30904262/>
16. Boom WH, Schaible UE, Achkar JM. The knowns and unknowns of latent *Mycobacterium tuberculosis* infection. *J Clin Invest*. [Internet]. 2021, Feb. [Cited 2021 May 20] 1;131(3):e136222. Available from: <https://pubmed.ncbi.nlm.nih.gov/33529162/>
17. Bussi C, Gutierrez MG. *Mycobacterium tuberculosis* infection of host cells in space and time. *FEMS Microbiol Rev*. [Internet]. 2019 Jul. [Cited 2021 May 20] 1;43(4):341-361. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6606852/>
18. Molina-Romero C, Arrieta O, Hernández-Pando R. Tuberculosis and lung cancer. *Salud Publica Mex*. [Internet]. 2019 May-Jun. [Cited 2021 Apr 13];61(3):286-291. Available from: <https://pubmed.ncbi.nlm.nih.gov/31276345/>
19. Corrales L, Rosell R, Cardona AF, Martín C, Zatarain-Barrón ZL, Arrieta O. Lung cancer in never smokers: The role of different risk factors other than tobacco smoking. *Crit Rev Oncol Hematol*. [Internet]. 2020 Apr [Cited 2021 Aug 25];148:102895. Available from: <https://pubmed.ncbi.nlm.nih.gov/32062313/>
20. Cukic V. The Association Between Lung Carcinoma and Tuberculosis. *Med Arch*.

- [Internet] 2017 Jun [Cited 2021 Sep 18] ;71(3):212-214. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5585804/>
21. Hong, S., Mok, Y., Jeon, C., Jee, S.H. and Samet, J.M. Tuberculosis, smoking and risk for lung cancer incidence and mortality. *Int. J. Cancer*. [Internet] 2016 [Cited 2021 Sep 19]; 139: 2447-2455. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/ijc.30384>
 22. Hwang SY, Kim JY, Lee HS, Lee S, Kim D, Kim S, et al. Pulmonary tuberculosis and risk of lung cancer: A systematic review and meta-analysis. *J Clin Med* [Internet]. 2022 [Cited 2021 Sep 25];11(3):765. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8836400/#>
 23. Castro-Garza J, Luévano-Martínez ML, Villarreal-Treviño L, Gosálvez J, Fernández JL, Dávila-Rodríguez MI, García-Vielma C, González-Hernández S, Cortés-Gutiérrez EI. Mycobacterium tuberculosis promotes genomic instability in macrophages. *Mem Inst Oswaldo Cruz*. [Internet]. 2018 Mar. [Cited 2021 11 Apr];113(3):161-166. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5804307/>
 24. Lochab S, Singh Y, Sengupta S, Nandicoori VK. Mycobacterium tuberculosis exploits host ATM kinase for survival advantage through SecA2 secretome. *Elife*. [Internet]. 2020, Mar. [Cited 2021 12 Apr] 30;9:e51466 Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7162654/>
 25. Bespyatykh JA, Shitikov EA, Ilina EN. Proteomics for the Investigation of Mycobacteria. *Acta Naturae*. [Internet]. 2017 Jan-Mar. [Cited 2021 May 20];9(1):15-25. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5406656/>
 26. Schubert OT, Aebersold R. Microbial Proteome Profiling and Systems Biology: Applications to Mycobacterium tuberculosis. *Adv Exp Med Biol*. [Internet]. 2015. [Cited 2021 May 20]; 883:235-54. Available from: <https://pubmed.ncbi.nlm.nih.gov/26621471/>
 27. Tucci P, Portela M, Chetto CR, González-Sapienza G, Marín M. Integrative proteomic and glycoproteomic profiling of Mycobacterium tuberculosis culture filtrate. *PLoS One*. [Internet]. 2020 Mar. [Cited 2021 May 21] 3;15(3): e0221837. Available from: <https://pubmed.ncbi.nlm.nih.gov/32126063/>
 28. Kelkar DS, Kumar D, Kumar P, Balakrishnan L, et al. Proteogenomic analysis of Mycobacterium tuberculosis by high resolution mass spectrometry. *Mol Cell Proteomics*. [Internet]. 2011 Dec. [Cited 2021 Aug 30] ;10(12):M111.011627. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3275902/>
 29. Bell C, Smith GT, Sweredoski MJ, Hess S. Characterization of the Mycobacterium

- tuberculosis proteome by liquid chromatography mass spectrometry-based proteomics techniques: a comprehensive resource for tuberculosis research. *J Proteome Res.* [Internet]. 2012 Jan. [Cited 2021 Jun 21] 1;11(1):119-30. Available from: <https://pubmed.ncbi.nlm.nih.gov/22053987/>
30. Peters JS, Calder B, Gonnelli G, Degroeve S, Rajaonarifara E, Mulder N, Soares NC, Martens L, Blackburn JM. Identification of Quantitative Proteomic Differences between *Mycobacterium tuberculosis* Lineages with Altered Virulence. *Front Microbiol.* [Internet] 2016 May. [Cited 2021 Aug 28] 31;7:813. Available from: <https://www.frontiersin.org/articles/10.3389/fmicb.2016.00813/full>
 31. Gopinath V, Raghunandan S, Gomez RL, Jose L, Surendran A, Ramachandran R, Pushparajan AR, Mundayoor S, Jaleel A, Kumar RA. Perfilado del proteoma de *Mycobacterium tuberculosis* durante la latencia y la reactivación. *Proteómica de células mol.* [Internet]. Agosto de 2015. [Cited 2021 Aug 20]; 14 (8): 2160-76. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4528245/>
 32. Albrethsen J, Agner J, Piersma SR, Højrup P, Pham TV, Weldingh K, Jimenez CR, Andersen P, Rosenkrands I. El perfil proteómico de *Mycobacterium tuberculosis* identifica sistemas de toxina-antitoxina que responden a la falta de nutrientes. *Proteómica de células mol.* [Internet]. Mayo de 2013. [Cited 2021 Aug 20]; 12 (5): 1180-91. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3650330/>
 33. Putim C, Phaonakrop N, Jaresitthikunchai J, Gamngoen R, Tragoolpua K, Intorasoot S, et al. Secretome profile analysis of multidrug-resistant, monodrug-resistant and drug-susceptible *Mycobacterium tuberculosis*. *Arch Microbiol.* [Internet] 2017 Mar. [Cited 2021 Jun 17];200(2):299-309. Available from: <https://pubmed.ncbi.nlm.nih.gov/29119205/>
 34. Khan S, Imran A, Khan AA, Abul Kalam M, Alshamsan A. Systems Biology Approaches for the Prediction of Possible Role of *Chlamydia pneumoniae* Proteins in the Etiology of Lung Cancer. *PLoS One.* [Internet]. 2016 Feb. [Cited 2021 May 22]. 12;11(2): e0148530. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4752481/>
 35. Li J, Zakariah M, Malik A, Ola MS, Syed R, Chaudhary AA, Khan S. Analysis of *Salmonella typhimurium* Protein-Targeting in the Nucleus of Host Cells and the Implications in Colon Cancer: An in-silico Approach. *Infect Drug Resist.* [Internet]. 2020 Jul. [Cited 2021 May 22]. 20;13:2433-2442. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7381790/>
 36. Khan S, Zaidi S, Alouffi AS, Hassan I, Imran A, Khan RA. Computational Proteome-Wide

- Study for the Prediction of Escherichia coli Protein Targeting in Host Cell Organelles and Their Implication in Development of Colon Cancer. ACS Omega. [Internet]. 2020 Mar. [Cited 2021 May 24]. 30;5(13):7254- 7261. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7143404/>
37. Wang Y, Imran A, Shami A, Chaudhary AA, Khan S. Decipher the Helicobacter pylori Protein Targeting in the Nucleus of Host Cell and their Implications in Gallbladder Cancer: An insilico approach. J Cancer. [Internet]. 2021 Oct 25 [Cited 2021 Jun 20]; 12 (23): 7214-7222. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/pmid/34729122/>
 38. Khan S, Zakariah M, Palaniappan S. Computational prediction of Mycoplasma hominis proteins targeting in nucleus of host cell and their implication in prostate cancer etiology. Tumour Biol. [Internet] 2016 Aug [Cited 2021 Jun 20]; 37(8): 10805-13. Available from: <https://pubmed.ncbi.nlm.nih.gov/26874727/>
 39. Khan AA, Khan Z, Malik A, Shrivastava A, Jain SK, Alshamsan A. Computational prediction of Escherichia coli proteins host subcellular targeting and their implications in colorectal cancer etiology. Cancer Lett. [Internet] 2015 Aug 1 [Cited 2021 Jun 20];364(1):25-32. Available from: <https://pubmed.ncbi.nlm.nih.gov/25917565/>
 40. Scanu T, Spaapen Robbert M, Bakker Jeroen M, Pratap C, Wu L, Hofland I, et al. Salmonella Manipulation of Host Signaling Pathways Provokes Cellular Transformation Associated with Gallbladder Carcinoma. Cell Host & Microbe [Internet]. 2015 Jun [cited 2022 Nov 15];17(6):763–74. Available from: <https://pubmed.ncbi.nlm.nih.gov/26028364/>
 41. Hyuna Sun, Jacques Ferlay, Rebecca L. Siegel, Mathieu Laversanne ,Isabelle Soerjomataram ,Ahmedin Jemal ,Freddie Bray. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. Cancer Journal for Clinicians. [Internet] February 2021 [Cited 2022 Jun 20]; 71(3):209-249. Available from: <https://acsjournals.onlinelibrary.wiley.com/doi/full/10.3322/caac.21660>
 42. Master W. Vigilancia de tuberculosis [Internet]. CDC MINSA. CDC MINSA; 2020 [cited 2023 Mar 1]. Available from: <https://www.dge.gob.pe/portalnuevo/vigilancia-epidemiologica/vigilancia-de-tuberculosis/>
 43. Bateman A, Martin M-J, Orchard S, Magrane M, Agivetova R, Ahmad S, et al. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Research [Internet]. 2020 Nov 25 [cited 2021 Sep 28];49(D1):D480–9. Available from: <https://academic.oup.com/nar/article/49/D1/D480/6006196>
 44. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Ostell J, Pruitt KD, Sayers EW.

- GenBank. *Nucleic Acids Res.* [Internet] 2018 Jan [cited 2022 Sept 25] 46(D1):D41-D47. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5753231/>
45. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology* [Internet]. 2019 Feb 18 [cited 2021 Sep 28];37(4):420–3. Available from: <https://pubmed.ncbi.nlm.nih.gov/30778233/>
 46. Bendtsen JD, Jensen LJ, Blom N, von Heijne G, Brunak S. Feature-based prediction of non-classical and leaderless protein secretion. *Protein Engineering Design and Selection* [Internet]. 2004 May 4 [cited 2021 Sep 28];17(4):349–56. Available from: <https://academic.oup.com/peds/article/17/4/349/1672190>
 47. Wan S, Mak M-W, Kung S-Y. FUEL-mLoc: feature-unified prediction and explanation of multi-localization of cellular proteins in multiple organisms. *Bioinformatics* [Internet]. 2016 Dec 23 [cited 2021 Sep 28];btw717. Available from: <https://pubmed.ncbi.nlm.nih.gov/28011780/>
 48. Pierleoni A, Martelli PL, Fariselli P, Casadio R. BaCellLo: a balanced subcellular localization predictor. *Bioinformatics* [Internet]. 2006 Jul 15 [cited 2021 Sep 28];22(14):e408–16. Available from: <https://academic.oup.com/bioinformatics/article/22/14/e408/228072?login=true>
 49. Yu C-S, Lin C-J, Hwang J-K. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based onn-peptide compositions. *Protein Science* [Internet]. 2004 May [cited 2021 Sep 28];13(5):1402–6. Available from: <https://pubmed.ncbi.nlm.nih.gov/15096640/>
 50. Goldberg T, Hecht M, Hamp T, et al. LocTree3 prediction of localization. *Nucleic Acids Res.* [Internet]. 2014 [cited 2021 Sep 28];42 (Web Server issue). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4086075/>
 51. Garg VK, Avashthi H, Tiwari A, Jain PA, Ramkete PWR, Kayastha AM, et al. MFPPi – Multi FASTA ProtParam Interface. *Bioinformatics* [Internet]. 2016 Apr 10 [cited 2021 Sep 28]; 12(2): 74–7. Available from: <https://pubmed.ncbi.nlm.nih.gov/28104964/>
 52. Cautain B, Hill R, de Pedro N, Link W. Components and regulation of nuclear transport processes. *FEBS J.* [Internet]. 2015 Feb. [cited 2021 Dec 4];282(3):445-62. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7163960/>
 53. Carbon S, Douglass E, Good BM, Unni DR, Harris NL, Mungall CJ, et al. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Research* [Internet]. 2020 Dec 8 [cited 2021 Sep 28]; 49 (D1): D325–34. Available from:

<https://pubmed.ncbi.nlm.nih.gov/33290552/>

54. Karp PD, Billington R, Caspi R, et al. The BioCyc collection of microbial genomes and metabolic pathways. *Brief Bioinform* [Internet]. 2019 [cited 2021 Ago 20] ; 20 (4): 1085-1093. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6781571/>
55. Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics*. 2009 [cited 2021 Ago 26]; 25 (22): 3045-3046. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2773257/>
56. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar G, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. *Nucleic Acids Research* [Internet]. 2020 Oct 30 [cited 2021 Sep 28]; 49 (D1): D412–9. Available from: <https://academic.oup.com/nar/article/49/D1/D412/5943818>
57. Blum M, Chang H-Y, Chuguransky S, Grego T, Kandasamy S, Mitchell A, et al. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research* [Internet]. 2020 Nov 6 [cited 2021 Sep 28]; 49 (D1): D344–54. Available from: <https://academic.oup.com/nar/article/49/D1/D344/5958491>
58. Liu R, Hu J. DNABind: a hybrid algorithm for structure-based prediction of DNA-binding residues by combining machine learning- and template-based approaches. *Proteins*. [Internet] 2013 Nov [cited 2021 Sep 28];81(11):1885-99. Available from: <https://pubmed.ncbi.nlm.nih.gov/23737141/>
59. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Research* [Internet]. 2008 May 19 [cited 2021 Sep 28];36(Web Server):W5–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/18440982/>
60. wwPDB consortium. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res* [Internet]. 2019 Jan [cited 2022 Apr 15]. 8;47(D1):D520-D528. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6324056/>
61. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. [Internet] 2021 [cited 2021 Oct 17]; 596 (7873) : 583-589. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8371605/>
62. Shuvo MH, Gulfam M, Bhattacharya D. DeepRefiner: high-accuracy protein structure refinement by deep network calibration. *Nucleic Acids Res* [Internet]. 2021 [cited 2021 Oct 17]; 49 (W1): W147-W152. Available from:

- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8262753/>
63. Pettersen EF, Goddard TD, Huang CC, et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.*[Internet] 2021 [cited 2021 Oct 25]; 30 (1): 70-82. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7737788/>
 64. Yan Y, Huang SY. Modeling Protein-Protein or Protein-DNA/RNA Complexes Using the HDOCK Webserver. *Methods Mol Biol.* [Internet] 2020 [cited 2021 Apr 25]; 2165: 217-229. Available from: <https://pubmed.ncbi.nlm.nih.gov/32621227/>
 65. Melissa F Adasme, Katja L Linnemann, Sarah Naomi Bolz, Florian Kaiser, Sebastian Salentin, V Joachim Haupt, Michael Schroeder, PLIP 2021: expanding the scope of the protein–ligand interaction profiler to DNA and RNA, *Nucleic Acids Research.* [Internet] 2 July 2021 [cited 2021 Sept 20] 49:530–534. Available from: <https://doi.org/10.1093/nar/gkab294>
 66. Tokmakov AA, Kurotani A, Sato K-I. Protein pI and Intracellular Localization. *Frontiers in Molecular Biosciences* [Internet]. 2021 Nov 29 [cited 2023 Jan 22];8. Available from: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.775736/full#B35>
 67. Data P. RCSB PDB - 7KGB: CryoEM structure of A2296-methylated Mycobacterium tuberculosis ribosome bound with SEQ-9 [Internet]. Rcsb.org. 2020 [cited 2022 Sep 14]. Available from: <https://www.rcsb.org/structure/7KGB>
 68. Kumari M, Pal RK, Mishra AK, Tripathi S, Biswal BK, Srivastava KK, et al. Structural and functional characterization of the transcriptional regulator Rv3488 of Mycobacterium tuberculosis H37Rv. *Biochemical Journal* [Internet]. 2018 Nov 9 [cited 2021 Dec 6];475(21):3393–416. Available from: <https://portlandpress.com/biochemj/article-abstract/475/21/3393/49939/Structural-and-functional-characterization-of-the?redirectedFrom=fulltext>
 69. Teplova M, Malinina L, Darnell JC, Song J, Lu M, Abagyan R, Musunuru K, Teplov A, Burley SK, Darnell RB, Patel DJ. Protein-RNA and protein-protein recognition by dual KH1/2 domains of the neuronal splicing factor Nova-1. *Structure* [Internet]. 2011 Jul 13;19(7):930-44 [cited 2022 Apr 6]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3134789/>
 70. Park SC, Kwak YM, Song WS, Hong M, Yoon SI. Structural basis of effector and operator recognition by the phenolic acid-responsive transcriptional regulator PadR. *Nucleic Acids Res* [Internet]. 2017 Dec 15 [cited 2022 Apr 6];45(22):13080-13093. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5728393/>
 71. Gilboa R, Zharkov DO, Golan G, Fernandes AS, Gerchman SE, Matz E, Kycia JH,

- Grollman AP, Shoham G. Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA. *J Biol Chem* [Internet]. 2002 May 31 [cited 2022 Apr 6];277(22):19811-6. Available from: [https://www.jbc.org/article/S0021-9258\(20\)85048-0/fulltext](https://www.jbc.org/article/S0021-9258(20)85048-0/fulltext)
72. Kim JM, Choe MH, Asaithambi K, Song JY, Lee YS, Lee JC, Seo JH, Kang HL, Lee KH, Lee WK, Cho MJ, Rhlee KH, Youn HS, Baik SC. Helicobacter pylori HP0425 Targets the Nucleus with DNase I-Like Activity. *Helicobacter*. 2015 Jun;21(3):218-25. Available from: <https://pubmed.ncbi.nlm.nih.gov/26395879/>
73. Lebreton A, Lakisic G, Job V, Fritsch L, Tham TN, Camejo A, et al. A Bacterial Protein Targets the BAHD1 Chromatin Complex to Stimulate Type III Interferon Response. *Science* [Internet]. 2011 Mar 11 [cited 2023 Jan 20];331(6022):1319–21. Available from: <https://pubmed.ncbi.nlm.nih.gov/21252314/>
74. Le LHM, Ying L, Ferrero RL. Nuclear trafficking of bacterial effector proteins. *Cellular Microbiology* [Internet]. 2021 Mar 2 [cited 2021 Dec 6];23(6). Available from: <https://onlinelibrary.wiley.com/doi/10.1111/cmi.13320?af=R>
75. Feltcher ME, Sullivan JT, Braunstein M. Protein export systems of Mycobacterium tuberculosis: novel targets for drug development? *Future Microbiol*. 2010 Oct;5(10):1581-97. Available from: [https://pubmed.ncbi.nlm.nih.gov/21073315/#:~:text=tuberculosis%20also%20has%20specialized%20export,processes%20\(ESX%2D3\).](https://pubmed.ncbi.nlm.nih.gov/21073315/#:~:text=tuberculosis%20also%20has%20specialized%20export,processes%20(ESX%2D3).)
76. Miller BK, Zulauf KE, Braunstein M. The Sec Pathways and Exportomes of Mycobacterium tuberculosis. *Microbiol Spectr*. [Internet]. 2017 Apr [cited 2021 Dec 3];5(2). Available from: <https://journals.asm.org/doi/10.1128/microbiolspec.TB2-0013-2016>
77. Bendtsen JD, Kiemer L, Fausbøll A, Brunak S. Non-classical protein secretion in bacteria. *BMC Microbiology* [Internet]. 2005 Oct 7 [cited 2021 Dec 4];5(1). Available from: <https://bmcmicrobiol.biomedcentral.com/articles/10.1186/1471-2180-5-58>
78. Martin RM, Ter-Avetisyan G, Herce HD, Ludwig AK, Lättig-Tünnemann G, Cardoso MC. Principles of protein targeting to the nucleolus. *Nucleus* [Internet]. 2015 Jul 4 [cited 2023 Jan 24];6(4):314–25. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4615656/>
79. Lee JH, Jun SH, Kim J-M, Baik SC, Lee JC. Morphological changes in human gastric epithelial cells induced by nuclear targeting of Helicobacter pylori urease subunit A. *Journal of Microbiology* [Internet]. 2015 May 30 [cited 2021 Dec 4];53(6):406–14.

Available from: <https://link.springer.com/article/10.1007/s12275-015-5085-5>

80. D'Souza D, Das BC. Genotoxic effects of Mycobacterium leprae infection in humans. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis [Internet]. 1994 Mar [cited 2023 Jan 24];305(2):211–22. Available from: <https://www.sciencedirect.com/science/article/abs/pii/0027510794902410?via%3Dihub>
81. Feng S, Liu Y, Liang W, El-Sayed Ahmed MAE-G, Zhao Z, Shen C, et al. Involvement of Transcription Elongation Factor GreA in Mycobacterium Viability, Antibiotic Susceptibility, and Intracellular Fitness. Frontiers in Microbiology [Internet]. 2020 Mar 23 [cited 2023 Jan 23];11. Available from: <https://www.frontiersin.org/articles/10.3389/fmicb.2020.00413/full>
82. Patil AG, Sang PB, Govindan A, Varshney U. Mycobacterium tuberculosis MutT1 (Rv2985) and ADPRase (Rv1700) proteins constitute a two-stage mechanism of 8-oxo-dGTP and 8-oxo-GTP detoxification and adenosine to cytidine mutation avoidance. J Biol Chem.[Internet] 2013 Apr 19 [cited 2023 Jan 20];288(16):11252-62. Available from: <https://pubmed.ncbi.nlm.nih.gov/23463507/>
83. Urlaub H;Kruft V;Bischof O;Müller EC;Wittmann-Liebold B. Protein-rRNA binding features and their structural and functional implications in ribosomes as determined by cross-linking studies. The EMBO journal [Internet]. 1995 [cited 2021 Dec 5];14(18). Available from: <https://pubmed.ncbi.nlm.nih.gov/7556101/>
84. Valverde R, Edwards L, Regan L. Structure and function of KH domains. FEBS J [Internet]. 2008 Jun[cited 2021 Dec 28];275(11):2712-26 : Available from:<https://pubmed.ncbi.nlm.nih.gov/18422648/>
85. Yadav M, Singh RS, Hogan D, Vidhyasagar V, Yang S, Chung IYW, et al. The KH domain facilitates the substrate specificity and unwinding processivity of DDX43 helicase. Journal of Biological Chemistry [Internet]. 2021 Jan [cited 2021 Dec 6];296:100085. Available from: [https://www.jbc.org/article/S0021-9258\(20\)00073-3/fulltext](https://www.jbc.org/article/S0021-9258(20)00073-3/fulltext)
86. Ruggero D. Translational Control in Cancer Etiology. Cold Spring Harbor Perspectives in Biology [Internet]. 2012 Jul 5 [cited 2022 Sep 22];5(2):a012336–6. Available from: <https://pubmed.ncbi.nlm.nih.gov/22767671/>
87. Graifer D, Malygin A, Zharkov DO, Karpova G. Eukaryotic ribosomal protein S3: A constituent of translational machinery and an extraribosomal player in various cellular processes. Biochimie [Internet]. 2014 Apr [cited 2022 Sep 15];99:8–18. Available from: <https://www.sciencedirect.com/science/article/abs/pii/S0300908413003970>
88. Kechavarzi B, Janga S. Dissecting the expression landscape of RNA-binding proteins in

- human cancers. *Genome Biology* [Internet]. 2014 [cited 2022 Sep 10];15(1):R14. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4053825/>
89. Isom CE, Menon SK, Thomas LM, West AH, Richter-Addo GB, Karr EA. Crystal structure and DNA binding activity of a PadR family transcription regulator from hypervirulent *Clostridium difficile* R20291. *BMC Microbiol.* [Internet] 2016 Oct 4 [Cited 2022 Dec 3];16(1):231. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5050560/>
90. Lee TI, Young RA. Transcriptional regulation and its misregulation in disease. *Cell.* [Internet] 2013 Mar 14 [cited 2022 Aug 6]; 152(6): 1237-51. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3640494/>
91. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. *Nat Med.* [Internet] 2004 Aug [cited 2022 Aug 20]; 10(8): 789-99. Available from: <https://pubmed.ncbi.nlm.nih.gov/15286780/>
92. Prakash A, Doublie S, Wallace SS. The Fpg/Nei Family of DNA Glycosylases. *Progress in Molecular Biology and Translational Science* [Internet]. 2012 [cited 2021 Dec 6];71–91. Available from: <https://www.sciencedirect.com/science/article/abs/pii/B9780123876652000043>
93. Sugahara M, Mikawa T, Kumasaka T, Yamamoto M, Kato R, Fukuyama K, et al. Crystal structure of a repair enzyme of oxidatively damaged DNA, MutM (Fpg), from an extreme thermophile, *Thermus thermophilus* HB8. *The EMBO Journal* [Internet]. 2000 Aug 1 [cited 2021 Dec 6];19(15):3857–69. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC306600/>
94. Guo Y, Bandaru V, Jaruga P, Zhao X, Burrows CJ, Iwai S, Dizdaroglu M, Bond JP, Wallace SS. The oxidative DNA glycosylases of *Mycobacterium tuberculosis* exhibit different substrate preferences from their *Escherichia coli* counterparts. *DNA Repair (Amst).* [Internet] 2010 Feb 4 [cited 2022 Jul 6]; 9(2):177-90. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2836239/>
95. Jacobs AL, Schär P. DNA glycosylases: in DNA repair and beyond. *Chromosoma.* [cited 2022 Jul 22] 2012 Feb [cited 2022 Jul 6];121 (1): 1-20. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3260424/>

ANEXOS

Anexo 1: Lista integrada del proteoma de *M. tuberculosis* (n=3753)

Disponibile en el siguiente Link:

<https://docs.google.com/spreadsheets/d/1vG4TGPlcL6nB09cqjwf2ZZhgAx6QZEetTGL5Mi9iumY/edit#gid=0>

Anexo 2. Proteínas del secretoma de *M. tuberculosis* determinadas experimentalmente (n=133) por Putim C. (2017), que contaban con secuencias en Uniprot y Genbank.

ID uniprot	Nombre de la proteína	Nombre del gen	Tamaño (aa)
P9WK09	Metalotioneína, MT, metalotioneína de unión a cobre, metalotioneína de unión a Cu(I), metalotioneína micobacteriana	mymT, Rv0186A	53
I6XFT2	Proteína no caracterizada	Rv2970A	56
A0A089QKZ7	Proteína no caracterizada Rv1155A	Rv1155A, LH57_06350	62
I6X8G2	Proteína conservada	Rv0078B	68
P9WKY5	Histona acetiltransferasa, HAT	Rv3423.1	76
O50456	Antitoxina VapB33	vapB33, Rv1241	86
O53353	Regulador transcripcional WhiB2, Probable acompañante WhiB2	whiB2, Rv3260c	89
P71926	Proteína no caracterizada	Rv2422	90
P9WKN9	Proteína no caracterizada Rv0942	Rv0942	92
I6X642	Proteína no caracterizada	Rv3047c	94
Q6MX18	Proteína similar a ESAT-6 EsxS	esxS, Rv3020c, LH57_16495	97
P9WHB9	50S proteína ribosomal L23	rpIW, Rv0703, MTCY210.22	100
B5TV88	Proteína similar a ESAT-6	cfp-10	100
P9WNK5	Proteína similar a ESAT-6 EsxB, antígeno filtrado de cultivo de 10 kDa CFP-10, CFP-10, proteína antigénica secretada MTSA-10	esxB, cfp10, lhp, mtsA10, Rv3874, MTV027.09	100
P9WMH7	Regulador transcripcional ClgR, regulador similar a ClpR, regulador del gen clp	CLGR, Rv2745c	112
O50392	Proteína no caracterizada	Rv3363c	122
O53173	Proteína de membrana integral probablemente conservada	Rv2446c	123
P95144	Proteína conservada	Rv1871c	129
P96238	Posible proteína reguladora de la transcripción	Rv3840	137
P9WLL7	Biliverdina reductasa dependiente de F420H (2), F-BVR	Rv2074, MTCY49.13	137

O53839	Posible desaminasa	Rv0828c	140
P9WFC9	Proteína de estrés universal Rv1636, USP Rv1636	Rv1636, TB15.3	146
P9WN87	Regulador transcripcional FurA	furA, Rv1909c, MTCY180.09	147
I6X666	Proteína no caracterizada	Rv3076	158
I6XEX3	Posible proteasa del profago PhiRv2	Rv2651c	177
P9WHE7	50S proteína ribosomal L10	rpIJ, Rv0651, MTCY20H10.32	178
I6XW93	Proteína no caracterizada	Rv0762c	181
I6YAY5	Proteína no caracterizada	Rv3033	182
P9WFL1	UPF0232 proteína Rv0004	Rv0004, MTCY10H4.02	187
P9WNM3	Factor de elongación P, EF-P	EFP, Rv2534c, MTCY159.22	187
P9WPC5	Subunidad proteolítica 1 de proteasa Clp dependiente de ATP, Endopeptidasa Clp 1	clpP1, clpP, Rv2461c, MTV008.17c	200
P9WKK5	Isopentenil-difosfato Delta-isomerasa, IPP isomerasa, IPP: DMAPP isomerasa, Isopentenil pirofosfato isomerasa	Idi, Rv1745c, MTCY04C12.29c, MTCY28.08c	203
P9WMR7	Pirofosfatasa dITP/XTP, Purina NTP pirofosfatasa no canónica, Purina NTP pirofosfatasa no estándar, Nucleósido-trifosfato difosfatasa, Nucleósido-trifosfato pirofosfatasa, NTPasa	Rv1341, MTCY02B10.05, MTCY130.26	204
P9WPG3	Probable fosfatidilglicerofosfato sintasa, PGP sintasa, PGPS, CDP-diacilglicerol--glicerol-3-fosfato fosfatidiltransferasa	pgsA3, Rv2746c	209
P9WHM5	Fosforribosilglicinamida formiltransferasa, 5'-fosforribosilglicinamida transformilasa, GAR transformilasa, GART	Rv0956	215
P9WMF9	Activador transcripcional de unión al ADN DevR/DosR	devR, dosR, Rv3133c	217
P95075	Posible L-fuculosa fosfato aldolasa FucA (L-fuculosa-1-fosfato aldolasa)	fucA, Rv0727c	218
P9WK81	Lipoproteína putativa LppA	lppA, Rv2543	219
P9WMH3	Regulador transcripcional global activado por AMPc similar a la PCR, proteína receptora de AMPc, PCR, proteína reguladora de AMPc	crp, Rv3676	224
P71837	Proteína no caracterizada	Rv0784	228
P9WJB3	Las fluoroquinolonas exportan proteína permeasa Rv2686c	Rv2686c	252
P9WFS3	Proteína de membrana de la familia TVP38/TMEM64 Rv1491c	Rv1491c, MTCY277.13c	252
I6XZC4	Probable deshidrogenasa/reductasa de cadena corta	Rv1941	256

P9WGT1	3-alfa-(o 20-beta)-hidroxiesteroide deshidrogenasa, 3alfa, 20beta-hidroxiesteroide deshidrogenasa dependiente de NADH	fabG3, Rv2002, MTCY39.16c	260
P9WGU3	Proteína de unión a molibdato ModA, proteína de unión a molibdato/tungstato ModA	MODA, Rv1857, MTCY359.16c	261
P9WMX9	Glicosiltransferasa no caracterizada Rv1514c	Rv1514c, MTCY277.36c	262
P9WFD3	Proteína de estrés universal Rv3134c, USP Rv3134c	Rv3134c	268
P71540	Probable enoil-CoA hidratasa EchA7 (Enoilhidrasa) (Acil-CoA hidratasa insaturada) (Crotonasa)	echA7, Rv0971c	269
P9WPE1	Proteína similar a la subunidad beta de citrato liasa	citE, Rv2498c	273
P9WH37	Proteína ribosómica S3 30S	rpsC, Rv0707, MTCY210.26	274
O06830	Proteína que contiene el dominio PlsC	Rv1428c	275
O53622	Posible oxidorreductasa	Rv0077c	276
O07427	Monoacilglicerol lipasa, MGL	Rv0183, LH57_01015	279
P9WFD9	Proteína de estrés universal Rv2028c, USP Rv2028c	Rv2028c	279
P71734	Proteína conservada	Rv2409c	279
O53664	3-hidroxiacil-tioéster deshidratasa X, 3-hidroxiubutiril-CoA deshidratasa, Enoil-CoA hidratasa 2	htdX, Rv0241c	280
P9WNC3	Formamidopirimidina-DNA glicosilasa 1, Fapy-DNA glicosilasa 1, 3.2.2.23, DNA-(sitio apurínico o apirimidínico) liasa MutM 1, AP liasa MutM 1	fpg1, mutM, Rv2924c, MTCY338.13c	289
P9WFZ9	Putativa proteína permeasa transportadora de péptidos Rv1282c	Rv1282c, MTCY373.01c, MTCY3H3.01	291
P9WMX3	Galactofuranosiltransferasa GlfT1, GalTr 1, 2.4.1.287, Arabinogalactano galactosiltransferasa 1, Ramnopiranosil-N-acetilglucosaminil-difosfo-decaprenol beta-1,4/1,5-galactofuranosiltransferasa, UDP-Galf: alfa-3-L-ramnosil-alfa-D -GlcNAc-pirofosfato poliprenol, galactofuranosil transferasa	glfT1, Rv3782	304
I6YGV9	Probable proteína de membrana integral transportadora de dipéptidos Transportador ABC DppB	dppB, Rv3665c	308
P9WKU3	Proteína no caracterizada Rv0497	Rv0497, MTCY20G9.23	310
P9WND3	Metionil-tRNA formiltransferasa	fmt, Rv1406, MTCY21B4.23	312
P9WMP5	Ácido delta-aminolevulínico deshidratasa, ALAD, ALADH, 4.2.1.24, Porfobilinógeno sintasa	hemoB, Rv0512, MTCY20G10.02	329
P9WQP3	Diacilglicerol aciltransferasa/micoliltransferasa Ag85A, DGAT, Acil-CoA:diacilglicerol aciltransferasa, antígeno 85 complejo A, 85A, Ag85A, proteína A de unión a fibronectina, Fbps A	fbpA, mpt44, Rv3804c, MTV026.09c	338
P9WMU5	factor de gluconeogénesis putativo	Rv1422, MTCY21B4.40	342

P9WMK5	4-hidroxi-2-oxohexanoato aldolasa ácido 4-hidroxi-2-ceto-pentanoico aldolasa, 4-hidroxi-2-oxopentanoato aldolasa, 4-hidroxi-2-oxovalerato aldolasa, HOA	hsaF, Rv3534c	346
P71847	(3aS,4S,5R,7aS)-5-hidroxi-7a-metil-1-oxo-octahidro-1H-indeno-4-carboxil-CoA deshidrogenasa, 5OH-HIC-CoA deshidrogenasa	ipdC, Rv3553	355
P9WPX9	3-deshidroquinato sintasa, DHQS	aroB, Rv2538c, MTCY159.18	362
P9WP69	Subunidad 2 de citocromo c oxidasa, subunidad 2 de citocromo aa3, polipéptido II de citocromo c oxidasa, Mtb92	ctaC, Rv2200c, MTCY190.11c	363
P9WL23	Proteína no caracterizada Rv2915c	Rv2915c, MTCY338.03c	370
P9WKI5	Oxidoreductasa no caracterizada Rv3410c	guaB3, Rv3410c, MTCY78.18	375
P95229	Proteína no caracterizada	Rv0272c	377
P95217	Posible proteína reguladora de la transcripción	Rv0260c	381
O06159	Dihidrolipoillisina-residuo aciltransferasa componente del complejo alfa-cetoácido deshidrogenasa de cadena ramificada, 2.3.1.168, componente E2 del complejo alfa-cetoácido deshidrogenasa de cadena ramificada, BCKADH E2, dihidrolipoillisina-residuo (2-metilpropanoil) transferasa	bkdC, pdhC, Rv2495c	393
P9WHR1	Putativa metaloproteasa de zinc Rip3	rasgar3, Rv2625c	393
I6YC99	Proteína de la familia Mce Mce4A	mce4A, Rv3499c	400
P96831	Probable acil-CoA deshidrogenasa FadE2	fadE2, Rv0154c	407
O53298	Proteína que contiene el dominio DUF6036	Rv3081	412
P9WFFV5	Histidina--tRNA ligasa, histidil-tRNA sintetasa, HisRS	hisS, Rv2580c, MTCY227.21	423
P9WHF1	Proteína de biosíntesis de riboflavina RibBA, 3,4-dihidroxi-2-butanona 4-fosfato sintasa, DHBP sintasa, GTP ciclohidrolasa-2, 3.5.4.25, GTP ciclohidrolasa II	costillaBA, costillaA, costillaA2, Rv1415, MTCY21B4.33	425
P9WLD1	Proteína no caracterizada Rv2305	Rv2305, MTCY339.04c	429
P9WHN3	Adenilosuccinato sintetasa, AMPSase, AdSS, IMP--aspartato ligasa	purA, Rv0357c, MTCY13E10.19c	432
O53734	Proteína que contiene el dominio amino_oxidasa	Rv0449c	439
P9WPX1	Homoserina deshidrogenasa, HDH	hom, thrA, Rv1294, MTCY373.14	441
P9WM57	Proteína no caracterizada Rv1128c	Rv1128c, MTCY22G8.17c	451
O33340	Probable aldehído deshidrogenasa AldC	aldC, Rv2858c	455
O06312	Transportador de magnesio MgtE	mgtE, Rv0362	460

P9WJF1	Decaprenilfosforil-beta-D-ribosa oxidasa, Decaprenilfosfo-beta-D-ribofuranosa 2-deshidrogenasa, Decaprenilfosforil-beta-D-ribofuranosa 2'-epimerasa subunidad DprE1, Decaprenil-fosforribosa 2'-epimerasa subunidad 1, Decaprenilfosforil- beta-D-ribofuranosa 2'-oxidasa, decaprenilfosforil-beta-D-ribosa 2-epimerasa subunidad de flavoproteína, decaprenilfosforil-beta-D-ribofuranosa 2-oxidasa dependiente de FAD	dprE1, Rv3790	461
O53512	Fosfo-2-dehidro-3-desoxiheptonato aldolasa AroG, 3-desoxi-D-arabino-heptuloseonato 7-fosfato sintasa, DAHP sintasa, Fosfo-2-ceto-3-desoxiheptonato aldolasa	aroG, Rv2178c	462
P9WQF5	Subunidad grande de 3-isopropilmalato deshidratasa, 4.2.1.33, Alfa-IPM isomerasa, IPMI, isopropilmalato isomerasa	leuC, Rv2988c, MTV012.02c	473
I6Y8H4	Probable deshidrogenasa	Rv0697	479
P9WMY7	D-inositol 3-fosfato glicosiltransferasa, N-acetilglucosamina-inositol-fosfato N-mshA, Rv0486, acetilglucosaminiltransferasa, GlcNAc-Ins-P N-MTCY20G9.12, acetilglucosaminiltransferasa		480
P9WHJ9	Proteína de reparación de ADN RadA, proteína de migración de ramas RadA, proteína de reparación de ADN Sms	RadA, Rv3585, MTV024.03	480
P9WH43	Proteína ribosómica S1 30S	rpsA, Rv1630, MTCY01B2.22	481
O53762	Monooxigenasa probable	Rv0565c	486
I6Y0R5	Dihidrofolato sintasa/folilpoliglutamato sintasa, DHFS/FPGS, Folilpoli-gamma-glutamato sintasa, Tetrahidrofolilpoliglutamato sintasa	folC, Rv2447c, LH57_13380	487
P9WN19	Cadena pequeña de glutamato sintasa [NADPH], subunidad beta de glutamato sintasa, cadena beta de GLTS	gltD, Rv3858c	488
P9WJ95	Proteína inducida por isoniazida IniC	inic, Rv0343	493
P71980	Posible carboxilasa	Rv1722	494
P95148	Proteína conservada	Rv1867	494
O53325	Posible amidasa (aminohidrolasa)	Rv3175	495
O05896	Posible proteína de membrana integral de transporte catiónico de aminoácidos	Rv3253c	495
I6Y4U4	Posible oxidoreductasa	Rv0794c	499
P9WM61	Proteína no caracterizada Rv0104	Rv0104, MTCY251.23	504
P9WFU9	Lisina: ARNt ligasa 1, lisil-ARNt sintetasa 1, LysRS 1	lysS1, lysS, Rv3598c, MTCY07H7B.24	505

P71883	Probable proteína de extrusión de nitrito 1 NarK1 (facilitador de nitrito 1)	narK1, Rv2329c	512
I6YEE6	Posible hidrolasa	Rv2800	549
P9WP37	Citocromo bc1 complejo citocromo b subunidad, Citocromo bc1 reductasa complejo subunidad QcrB, Ubiquinol--citocromo c reductasa citocromo b subunidad	qcrB, Rv2196, MTCY190.07	549
P9WLJ3	Proteína no caracterizada Rv2100	Rv2100, MTCY49.40	550
P9WIS7	Dihidrolipoillisina-residuo acetiltransferasa componente del complejo piruvato deshidrogenasa, componente del complejo peroxinitrito reductasa/peroxidasa, componente de PNR/P, componente de dihidrolipoamida acetiltransferasa del complejo piruvato deshidrogenasa, componente E2 del complejo piruvato deshidrogenasa, componente E2 de PDH	dlaT, sucB, Rv2215, MTCY190.26	553
O05316	Proteína conservada	Rv1215c	561
O05305	Proteína que contiene dominio AAA	Rv1204c	562
Q6MX34	Probable uroporfirina-III C-metiltransferasa HemD (Uroporfirinógeno III metilasa) (Urogen III metilasa) (SUMT) (Urogen III metilasa) (UROM)	dobladillo, Rv0511	565
P96414	Probable acil-CoA deshidrogenasa FadE4	fadE4, Rv0231	568
P9WHJ1	Subunidad enzimática RecBCD RecD, subunidad Exonucleasa V RecD, subunidad ExoV RecD, subunidad Helicasa/nucleasa RecBCD RecD	RecD, Rv0629c	575
P9WQJ7	Importación de micobactina Proteína de unión a ATP/permeasa IrtB, transportador B regulado por hierro	irtB, Rv1349, MTCY02B10.13	579
I6YDI9	Proteína que contiene el dominio PlsC	por favor, Rv2483c	580
I6Y7V6	Probable acil-CoA ligasa FadD31 (Acil-CoA sintetasa) (Acil-CoA sintasa)	modaD31, Rv1925	620
P9WMJ9	Proteína chaperona DnaK, HSP70, Proteína de choque térmico de 70 kDa, Proteína de choque térmico de 70	dnaK, Rv0350, MTCY13E10.10	625
P9WQJ3	Transportador ABC de ácidos grasos Proteína de unión a ATP/permeasa	Rv1272c, MTCY50.10	631
P9WN33	Putativa asparagina sintetasa [hidrolizante de glutamina]	asnB, Rv2201, MTCY190.12	652
P9WJQ1	Proteína similar a la nitrato reductasa NarX	narX, Rv1736c	652
P9WLZ7	Proteína reguladora multidominio Rv1364c, factor anti-sigma-F Rv1364c, factor anti-SigF, proteína-serina/treonina fosfatasa, 3.1.3.16, regulador multidominio putativo de SigF, MursiF, serina/treonina-proteína quinasa	Rv1364c, MTCY02B10.28c	653
P9WG45	subunidad B de la ADN girasa	gyrB, Rv0005, MTCY10H4.03	675

P9WPT5	ATPasa tipo P exportadora de manganeso, ATPasa tipo P transportadora de cationes CtpC, ATPasa transportadora de metales Mta72	ctpC, mtaA, Rv3270, MTCY71.10	718
P9WK17	Malato sintasa G	glcB, Rv1837c, MTCY1A11.06	741
P9WMQ1	ADN helicasa dependiente de ATP UvrD1	uvrD1, ivrd, pcrA, Rv0949, MTCY10D7.25c	771
I6YGX2	Peptidoglicano glicosiltransferasa	ponA2, Rv3682	810
P9WM41	Proteína no caracterizada Rv1278	Rv1278, MTCY50.04c	875
P9WG49	ADN topoisomerasa 1, ADN topoisomerasa I, Omega-proteína, Enzima relajante, Swivelasa, Enzima destorcedora	parte superiorA, Rv3646c, MTCY15C10.06	934
O53166	Aconitato hidratasa A, ACN, Aconitasa, (2R,3S)-2-metilisocitrato deshidratasa, (2S,3R)-3-hidroxibutano-1,2,3-tricarboxilato deshidratasa, similar a proteína sensible al hierro, IRP -similar, probable 2-metil-cis-aconitato hidratasa, proteína de unión a ARN	acn, Rv1475c, RVBD_1475c, P425_01532	943
P9WNT7	Subunidad alfa de la ADN polimerasa III	ADNE1, ADNE, Rv1547, MTCY48.18c	1184
P9WGF3	Proteína de partición cromosómica Smc	smc, Rv2922c, MTCY338.11c	1205

Anexo 3. Proteínas del proteoma que presentan Péptido señal (PS), predichas de secreción mediante el programa SignalP (n=342).

Uniprot ID	Nombre de la proteína	SignalP				Predicción*
		SP(Sec/SPI)	TAT(Ta t/SPI)	LIPO(Se c/SPII)	OTHER	
P71750	Probable gamma-glutamyltranspeptidasa GgtB (Gamma-glutamyltransferasa) (Glutamyl transpeptidasa)	0.008706	0.000446	0.990563	0.000285	LIPO(Sec/SPII)
P96264	Probable lipoproteína aminopeptidasa LpqL, Leucina aminopeptidasa, Lipoproteína LpqL	0.007887	0.000797	0.988021	0.003296	LIPO(Sec/SPII)
A0A089QRB9	Micolipanoato sintasa, policétido sintasa similar al ácido micocerosico sintasa, PKS similar a MAS, policétido sintasa similar a los ácidos micolipanoico/micolipénico sintasa	0.025168	0.001363	0.846557	0.126913	LIPO(Sec/SPII)

I6Y231	Probable Mas asociado a la membrana de ácido micocerósico multifuncional sintasa	0.023937	0.00163	0.545726	0.428707	LIPO(Sec/SPII)
I6YAY5	Proteína no caracterizada	0.061979	0.018227	0.916945	0.002849	LIPO(Sec/SPII)
P9WK65	Transportador putativo de dimicocerosato de ftiocerol LppX , lipoproteína LppX	0.000694	0.009956	0.988921	0.000429	LIPO(Sec/SPII)
L0TBR2	Posible flavoproteína	0.103374	0.135242	0.705964	0.05542	LIPO(Sec/SPII)
O53291	Probable lipoproteína periplásmica de unión a dicitrato de FEIII FecB	0.004239	0.000375	0.99498	0.000406	LIPO(Sec/SPII)
P9WGE9	Superóxido dismutasa [Cu-Zn]	0.015306	0.055856	0.927771	0.001067	LIPO(Sec/SPII)
P9WGT7	Proteína de unión a fosfato PstS 3, PBP 3, PstS-3, Antígeno Ag88	0.002002	0.000361	0.997215	0.000422	LIPO(Sec/SPII)
P96257	Probable lipoproteína de unión a glutamina GlnH (GLNBP)	0.008779	0.014469	0.976481	0.000271	LIPO(Sec/SPII)
O06392	Posible proteína tiorredoxina (proteína de intercambio tiol-disulfuro)	0.005686	0.064071	0.929478	0.000765	LIPO(Sec/SPII)
P71744	Probable lipoproteína de unión a sulfato SubI	0.102275	0.038205	0.801169	0.05835	LIPO(Sec/SPII)
P9WKD3	Beta-lactamasa, beta-lactamasa de clase A de Ambler	0.021704	0.28071	0.696134	0.001452	LIPO(Sec/SPII)
L7N6B0	Beta-hexosaminidasa LpqI, Beta-N-acetilglucosaminidasa, Beta-N-acetilhexosaminidasa	0.007334	0.008602	0.983238	0.000827	LIPO(Sec/SPII)
P9WNF3	Glicolipoproteína de superficie celular MPT83, Lipoproteína p23	0.000796	0.000608	0.998221	0.000375	LIPO(Sec/SPII)
P71969	Posible proteasa secretada	0.222441	0.02371	0.748602	0.005246	LIPO(Sec/SPII)
P9WK55	Lipoproteína LprA	0.004978	0.002204	0.991895	0.000923	LIPO(Sec/SPII)
P9WK45	Proteína transportadora de lipoarabinomanano LprG, lipoproteína de 27 kDa, antígeno P27, lipoproteína LprG, transportador de glicolípidos triacilados LprG, proteína de	0.000941	0.000377	0.998427	0.000255	LIPO(Sec/SPII)

	transferencia de triacilglicéridos LprG					
O53901	Policétido sintasa similar al ácido micocerosico sintasa, PKS similar a MAS, 2.3.1.-, Policétido sintasa Pks5	0.013878	0.001673	0.690466	0.293983	LIPO(Sec/SPII)
P9WK63	Lipoproteína putativa LpqE	0.003664	0.000368	0.995557	0.000411	LIPO(Sec/SPII)
P9WGT9	Proteína de unión a fosfato PstS 2, PBP 2, PstS-2	0.003607	0.000494	0.995611	0.000287	LIPO(Sec/SPII)
I6Y293	Lipoproteína probablemente conservada LppZ	0.003073	0.000718	0.995853	0.000356	LIPO(Sec/SPII)
P9WK29	Proteína no caracterizada Rv1899c	0.27762	0.195059	0.309391	0.217931	LIPO(Sec/SPII)
O05592	Proteína no caracterizada	0.033688	0.007719	0.95716	0.001432	LIPO(Sec/SPII)
P9WHR3	Serina proteasa Hip1, Hidrolasa importante para la patogénesis 1, Serina hidrolasa Hip1	0.02744	0.038811	0.913277	0.020472	LIPO(Sec/SPII)
P9WK47	Transportador de glicolípidos diacilados putativo LprF, lipoproteína LprF	0.013934	0.035698	0.947167	0.003201	LIPO(Sec/SPII)
P9WPK3	Carbamoil-fosfato sintasa de cadena larga, Carbamoil-fosfato sintetasa cadena de amoníaco	0.032154	0.007569	0.730033	0.230244	LIPO(Sec/SPII)
P9WK61	Lipoproteína LpqH, antígeno de lipoproteína de 19 kDa, transportador putativo LpqH, p19	0.001097	0.000169	0.998542	0.000191	LIPO(Sec/SPII)
O86337	Posible lipoproteína LprC	0.003606	0.000396	0.995717	0.000281	LIPO(Sec/SPII)
I6X7X3	Lipoproteína probablemente conservada LpqG	0.001505	0.000237	0.997858	0.000399	LIPO(Sec/SPII)
O53780	Lipoproteína LpqN	0.000712	0.000089	0.999067	0.000132	LIPO(Sec/SPII)
P9WGU1	Proteína de unión a fosfato PstS 1, PBP 1, PhoS1, PstS-1, glicolipoproteína de 38 kDa, lipoproteína de 38 kDa, P38, antígeno Ag78, antígeno proteico B, Pab	0.001412	0.000166	0.998203	0.000219	LIPO(Sec/SPII)
P9WK71	Lipoproteína putativa LppO	0.135023	0.084706	0.779045	0.001225	LIPO(Sec/SPII)

P9WK59	Lipoproteína putativa LpqT	0.007916	0.001517	0.989633	0.000935	LIPO(Sec/SPII)
P9WK37	Lipoproteína LpqB	0.004965	0.000215	0.994589	0.00023	LIPO(Sec/SPII)
O05582	Proteína que contiene el dominio DUF5642	0.124558	0.281534	0.455367	0.13854	LIPO(Sec/SPII)
O69725	Posible osmoprotector (glicina betaína/carnitina/colina/L-prolina) que se une a la lipoproteína ProX	0.148707	0.001127	0.848988	0.001179	LIPO(Sec/SPII)
P9WGC1	Proteína no caracterizada Rv1339	0.015438	0.001839	0.970251	0.012472	LIPO(Sec/SPII)
P9WGU7	Probable proteína de unión a monoacil fosfatidilinositol tetramanósido LpqW	0.000985	0.007421	0.991261	0.000333	LIPO(Sec/SPII)
O53505	Proteína LppM, lipoproteína putativa LppM	0.088186	0.045612	0.864463	0.001739	LIPO(Sec/SPII)
O33363	GDSL lipasa Rv0518	0.227047	0.005597	0.566708	0.200649	LIPO(Sec/SPII)
I6YGI4	Posible lipoproteína conservada LppH	0.042624	0.000492	0.955925	0.000958	LIPO(Sec/SPII)
P9WGU5	Proteína no caracterizada Rv1280c	0.098598	0.055308	0.738941	0.107152	LIPO(Sec/SPII)
P9WLG7	Proteína no caracterizada Rv2240c	0.012307	0.039439	0.939247	0.009007	LIPO(Sec/SPII)
P9WFN3	Lipoproteína putativa LppC, proteína UPF0098 Rv1911c	0.000743	0.000529	0.998407	0.000321	LIPO(Sec/SPII)
I6WZ30	Proteína rica en treonina conservada	0.017762	0.049246	0.9318	0.001191	LIPO(Sec/SPII)
O07417	Posible lipoproteína LprK de la familia Mce (lipoproteína de la familia Mce Mce1E)	0.004772	0.001017	0.992688	0.001523	LIPO(Sec/SPII)
O50416	Lipoproteína probablemente conservada LpqD	0.001362	0.000572	0.997816	0.00025	LIPO(Sec/SPII)
P9WK53	Lipoproteína putativa LprB	0.002546	0.00025	0.997008	0.000195	LIPO(Sec/SPII)
P9WP69	Subunidad 2 de citocromo c oxidasa, subunidad 2 de citocromo aa3, polipéptido II de citocromo c oxidasa, Mtb92	0.073238	0.043361	0.861292	0.022108	LIPO(Sec/SPII)

P9WGU3	Proteína de unión a molibdato ModA, proteína de unión a molibdato/tungstato ModA	0.005091	0.003937	0.990658	0.000314	LIPO(Sec/SPII)
P96265	Posible lipoproteína peptidasa LpqM	0.016563	0.00144	0.972312	0.009684	LIPO(Sec/SPII)
P95291	Lipoproteína probablemente conservada	0.009597	0.027288	0.962611	0.000504	LIPO(Sec/SPII)
O53850	D-alanil-D-alanina dipeptidasa, D-Ala-D-Ala dipeptidasa	0.034982	0.000377	0.963161	0.00148	LIPO(Sec/SPII)
P9WGU9	Lipoproteína de unión a trehalosa LpqY	0.045599	0.0018	0.950711	0.00189	LIPO(Sec/SPII)
I6X811	Probable lipoproteína de unión a dipéptido periplásmico DppA	0.009735	0.00032	0.988729	0.001215	LIPO(Sec/SPII)
P9WK77	Lipoproteína putativa LppJ	0.009006	0.019461	0.968064	0.003468	LIPO(Sec/SPII)
I6XFA6	Lipoproteína probable LppU	0.009891	0.000232	0.989535	0.000342	LIPO(Sec/SPII)
O50459	Lipoproteína probable LpqZ	0.011862	0.000282	0.987573	0.000283	LIPO(Sec/SPII)
I6X9E2	Lipoproteína probablemente conservada LpqO	0.002318	0.011823	0.985509	0.00035	LIPO(Sec/SPII)
O53489	Lipoproteína probable LppI	0.002992	0.000164	0.996586	0.000257	LIPO(Sec/SPII)
O06237	Lipoproteína probablemente conservada LppL	0.003534	0.005974	0.990065	0.000428	LIPO(Sec/SPII)
P9WPA3	Defosfo-CoA quinasa, Defosfocoenzima A quinasa	0.058055	0.002954	0.492097	0.446894	LIPO(Sec/SPII)
P9WK75	Lipoproteína putativa LppK	0.000801	0.000628	0.998399	0.000173	LIPO(Sec/SPII)
O53846	Posible lipoproteína LpqQ	0.007702	0.000441	0.991341	0.000516	LIPO(Sec/SPII)
P71740	Lipoproteína probablemente conservada LppR	0.00469	0.000581	0.994201	0.000527	LIPO(Sec/SPII)
O06295	Probable lipoprotein LpqJ	0.001245	0.000201	0.998406	0.000148	LIPO(Sec/SPII)
O53246	Lipoproteína probablemente conservada LppY	0.002452	0.000472	0.996552	0.000524	LIPO(Sec/SPII)
P95207	Posible lipoproteína conservada LpqK	0.001536	0.00055	0.997576	0.000338	LIPO(Sec/SPII)

P9WK69	Lipoproteína putativa LppP	0.012924	0.007835	0.976855	0.002385	LIPO(Sec/SPII)
P9WK81	Lipoproteína putativa LppA	0.002444	0.003113	0.991554	0.002889	LIPO(Sec/SPII)
O06155	Lipoproteína probablemente conservada LpqF	0.062053	0.25991	0.677512	0.000526	LIPO(Sec/SPII)
P9WNF7	Monooxigenasa putativa que contiene FAD MymA	0.04896	0.01459	0.80737	0.12908	LIPO(Sec/SPII)
I6Y3P1	Lipoproteína LprN	0.003313	0.000336	0.995751	0.000600	LIPO(Sec/SPII)
P9WK49	Lipoproteína putativa LprE	0.003613	0.004742	0.990924	0.000720	LIPO(Sec/SPII)
I6XYM2	Lipoproteína probablemente conservada DsbF	0.003435	0.000441	0.995705	0.000419	LIPO(Sec/SPII)
O33224	Lipoproteína probable LpqX	0.001333	0.000211	0.998181	0.000275	LIPO(Sec/SPII)
P9WK39	Lipoproteína LprP no caracterizada	0.000643	0.000225	0.998675	0.000457	LIPO(Sec/SPII)
P9WK43	Lipoproteína putativa LprH	0.005123	0.001710	0.992413	0.000754	LIPO(Sec/SPII)
O07750	Probable lipoprotein LppE	0.006977	0.000822	0.991584	0.000617	LIPO(Sec/SPII)
I6Y9J2	L,D-transpeptidasa 2 , LDT 2, Ldt(Mt2)	0.009068	0.003819	0.984382	0.002732	LIPO(Sec/SPII)
O07421	Proteína asociada a Mce probablemente conservada	0.062037	0.270953	0.35512	0.311889	LIPO(Sec/SPII)
O53716	Proteína no caracterizada	0.001646	0.000259	0.997722	0.000373	LIPO(Sec/SPII)
P9WK67	Lipoproteína putativa LppW	0.009595	0.000357	0.989711	0.000337	LIPO(Sec/SPII)
I6Y2A3	Probable lipoprotein LpqA	0.04354	0.006024	0.947568	0.002869	LIPO(Sec/SPII)
P9WL77	Lipoproteína no caracterizada Rv2585c	0.115025	0.070882	0.812556	0.001537	LIPO(Sec/SPII)
L7N699	Probable ácido graso-CoA ligasa FadD34 (ácido graso-CoA sintetasa) (ácido graso-CoA sintetasa)	0.028226	0.002903	0.549841	0.41903	LIPO(Sec/SPII)

P71655	Lipoproteína probablemente conservada LppV	0.008857	0.000288	0.990473	0.000382	LIPO(Sec/SPII)
O33346	Posible lipoproteína de unión a penicilina	0.00167	0.008853	0.988958	0.000519	LIPO(Sec/SPII)
P96903	Posible lipoproteína esterasa LpqC	0.004187	0.000246	0.994409	0.001159	LIPO(Sec/SPII)
P9WM17	Proteína no caracterizada Rv1351	0.128765	0.041027	0.600469	0.229738	LIPO(Sec/SPII)
P9WLP3	Proteína no caracterizada Rv1995	0.013504	0.43951	0.524864	0.022122	LIPO(Sec/SPII)
P9WK79	Lipoproteína putativa LppB	0.003609	0.00407	0.989719	0.002601	LIPO(Sec/SPII)
P96378	Lipoproteína probablemente conservada LpqU	0.005826	0.004602	0.983351	0.00622	LIPO(Sec/SPII)
I6Y461	Posible lipoproteína LprL de la familia Mce (lipoproteína de la familia Mce Mce2E)	0.046097	0.007125	0.861993	0.084784	LIPO(Sec/SPII)
O53971	Posible lipoproteína de la familia Mce LprM (lipoproteína de la familia Mce Mce3E)	0.017139	0.000289	0.980557	0.002015	LIPO(Sec/SPII)
P9WMX7	Biosíntesis de PGL/p-HBAD glicosiltransferasa Rv2957	0.096767	0.003158	0.817456	0.082619	LIPO(Sec/SPII)
O53963	Lipoproteína probablemente conservada LppF	0.006701	0.000303	0.992621	0.000376	LIPO(Sec/SPII)
P9WG29	Factor promotor de la reanimación RpfB	0.247929	0.000709	0.745793	0.005568	LIPO(Sec/SPII)
P9WK73	Lipoproteína putativa LppN	0.00226	0.000227	0.997315	0.000198	LIPO(Sec/SPII)
P9WIQ7	Proteína repetitiva exportada, proteína de superficie celular PirG, EXP53	0.098233	0.41447	0.423211	0.064085	LIPO(Sec/SPII)
P9WGE5	Poliamina aminopropiltransferasa, aminopropiltransferasa, PAPT, espermidina sintasa, SPDS, SPDSY	0.040715	0.009808	0.484423	0.465054	LIPO(Sec/SPII)
O53859	Lipoproteína LpqS	0.035944	0.003316	0.950971	0.009768	LIPO(Sec/SPII)
P9WK41	Lipoproteína LprI, Glicolipoproteína LprI, Inhibidor de lisozima LprI	0.011003	0.000184	0.988385	0.000428	LIPO(Sec/SPII)

P9WJ77	Proteína de membrana externa Rv1973	0.0732	0.025091	0.735073	0.166636	LIPO(Sec/SPII)
P95237	Lipoproteína probablemente conservada LppQ	0.010377	0.054802	0.933577	0.001243	LIPO(Sec/SPII)
I6XVY0	Posible lipoproteína conservada LpqP	0.022146	0.001502	0.974913	0.001439	LIPO(Sec/SPII)
O33192	Lipoproteína putativa LprJ	0.313315	0.041253	0.623605	0.021826	LIPO(Sec/SPII)
P9WG65	Antígeno secretado soluble MPT53	0.854457	0.072142	0.003428	0.069973	SP(Sec/SPI)
O07222	Proteína conservada	0.962258	0.013814	0.016622	0.007305	SP(Sec/SPI)
O53945	Mycosin-5 , MycP5 proteasa	0.87541	0.118636	0.005499	0.000454	SP(Sec/SPI)
P9WQN9	Diacilglicerol aciltransferasa/micoliltransferasa Ag85C , DGAT, Acil-CoA:diacilglicerol aciltransferasa, antígeno 85 complejo C, 85C, Ag85C, proteína C de unión a fibronectina, Fbps C	0.480892	0.350708	0.013385	0.155014	SP(Sec/SPI)
P9WFJ5	UPF0603 proteína Rv2345	0.986134	0.008761	0.003711	0.001395	SP(Sec/SPI)
I6XF52	Posible proteína rica en alanina y glicina transmembrana conservada	0.68389	0.306138	0.004385	0.005587	SP(Sec/SPI)
O50430	Antígeno de células T de bajo peso molecular TB8.4	0.990088	0.006836	0.002273	0.000803	SP(Sec/SPI)
I6YGX2	Peptidoglicano glicosiltransferasa	0.419173	0.088317	0.292943	0.199568	SP(Sec/SPI)
P9WIN9	Proteína inmunogénica MPT64, Antígeno MPT64	0.889733	0.000381	0.109658	0.000228	SP(Sec/SPI)
P9WP43	Carboxilesterasa Culp1, CFP21, proteína tipo cutinasa 1, Culp1	0.95108	0.042793	0.004293	0.001834	SP(Sec/SPI)
O07175	Probable serine protease PepA (Serine proteinase) (MTB32A)	0.560439	0.435641	0.003008	0.000912	SP(Sec/SPI)
P9WPR9	Proteína no caracterizada Rv1488	0.579051	0.008344	0.008139	0.404466	SP(Sec/SPI)
P9WIP1	Proteína inmunogénica MPT63 , proteína extracelular inmunoprotectora de 16 kDa, antígeno MPT63	0.995492	0.003423	0.000573	0.000512	SP(Sec/SPI)
I6YGW9	Proteína conservada	0.467159	0.072667	0.035036	0.425138	SP(Sec/SPI)

O53695	Mycosin-3, MycP3 proteasa	0.97168	0.020257	0.006279	0.001784	SP(Sec/SPI)
P9WQN7	Antígeno MPT51/MPB51	0.896461	0.095018	0.005758	0.002763	SP(Sec/SPI)
P9WIR7	Proteína secretada rica en alanina y prolina Apa, glicoproteína de 45 kDa, antígeno de 45/47 kDa, antígeno MPT-32, FAP-B, proteína de unión a fibronectina, proteína inmunogénica MPT32	0.576571	0.403249	0.018819	0.001362	SP(Sec/SPI)
P9WP35	Subunidad del complejo citocromo c del citocromo bc1, Subunidad del complejo del citocromo bc1 reductasa QcrC, Ubiquinol--subunidad del citocromo c reductasa del citocromo c	0.674784	0.251842	0.011013	0.062362	SP(Sec/SPI)
P96409	Proteína de membrana probablemente conservada	0.611897	0.009293	0.166435	0.212374	SP(Sec/SPI)
P9WKL3	Proteína no caracterizada Rv0559c	0.998261	0.000808	0.000715	0.000216	SP(Sec/SPI)
P95202	Posible proteína secretada	0.504448	0.081812	0.0449	0.36884	SP(Sec/SPI)
O53740	Proteína conservada	0.789192	0.207027	0.002978	0.000803	SP(Sec/SPI)
I6YF08	Esterasa Rv3036c	0.969697	0.002907	0.023853	0.003543	SP(Sec/SPI)
P9WLX9	Proteína no caracterizada Rv1419	0.923875	0.039554	0.007147	0.029424	SP(Sec/SPI)
P9WNF5	Proteína inmunogénica MPT70	0.971613	0.025282	0.002418	0.000687	SP(Sec/SPI)
P9WIN7	Antígeno de bajo peso molecular MTB12, CFP-2, antígeno de proteína de bajo peso molecular 2	0.860953	0.069454	0.042725	0.026868	SP(Sec/SPI)
P96872	Proteína conservada (proteína relacionada con CPSA)	0.798042	0.02636	0.0141	0.161498	SP(Sec/SPI)
P95206	Posible proteína secretada	0.944365	0.016844	0.036873	0.001918	SP(Sec/SPI)
O07726	Proteína conservada	0.958401	0.010598	0.027609	0.003392	SP(Sec/SPI)
O53444	Posible glicosil hidrolasa	0.764292	0.118349	0.041137	0.076223	SP(Sec/SPI)
P9WFN5	proteína UPF0098 Rv1910c	0.506502	0.132754	0.202108	0.158636	SP(Sec/SPI)
P9WIB9	Corismato mutasa secretada, CM	0.775619	0.217031	0.001525	0.005824	SP(Sec/SPI)
P9WIM9	Antígeno de 28 kDa rico en prolina	0.746865	0.193884	0.048619	0.010632	SP(Sec/SPI)
P9WIH1	inmunomodulador de la familia PE PE15	0.423583	0.318848	0.021053	0.236516	SP(Sec/SPI)
I6X8R5	Proteína de unión a hemo Rv0203	0.590556	0.385519	0.016449	0.007476	SP(Sec/SPI)

P9WM15	Proteína no caracterizada Rv1352	0.91307	0.08473	0.001285	0.000915	SP(Sec/SPI)
O05458	Mycosin-2, MycP2 proteasa	0.686419	0.299753	0.012385	0.001442	SP(Sec/SPI)
P9WGP7	Proteína de la familia de las epimerasas Rv2216	0.563292	0.108853	0.039975	0.28788	SP(Sec/SPI)
O53531	Proteína conservada	0.500893	0.173874	0.023731	0.301501	SP(Sec/SPI)
I6Y4U9	Peroxidasa decolorante de colorantes, DyP	0.491396	0.010952	0.086074	0.411579	SP(Sec/SPI)
O05859	Proteína conservada	0.889156	0.02317	0.038581	0.049093	SP(Sec/SPI)
O05854	Proteína conservada rica en alanina y valina	0.531384	0.355716	0.014621	0.098279	SP(Sec/SPI)
O06624	Posible proteína Inv	0.959692	0.036676	0.002848	0.000784	SP(Sec/SPI)
P9WKE7	Homoserina quinasa, HK, HSK	0.58745	0.038208	0.186576	0.187766	SP(Sec/SPI)
P9WGA1	Proteína translocasa proteína independiente de sec TATA	0.419375	0.009935	0.257195	0.313495	SP(Sec/SPI)
O07236	Posible proteína exportada conservada	0.744171	0.008506	0.217592	0.02973	SP(Sec/SPI)
O69721	Prefenato deshidrogenasa, PDH	0.779128	0.106283	0.028985	0.085604	SP(Sec/SPI)
O53527	Proteína no caracterizada	0.968525	0.029957	0.0011	0.000418	SP(Sec/SPI)
O53361	Fosfatidilinositol-3-fosfatasa, fosfatasa PI3P, fosfatasa ácida SapM, fosfatasa ácida secretada	0.676207	0.089299	0.02443	0.210065	SP(Sec/SPI)
P9WJS7	Proteína accesoria de exportación de sideróforos MmpS5	0.478685	0.004101	0.120499	0.396715	SP(Sec/SPI)
I6X7P2	Proteína no caracterizada	0.987128	0.006864	0.005325	0.000682	SP(Sec/SPI)
I6YGW2	posible proteasa	0.869071	0.126759	0.003485	0.000685	SP(Sec/SPI)
P71965	Posible proteína rica en alanina y valina exportada	0.987879	0.001558	0.010089	0.000474	SP(Sec/SPI)
I6XI06	Proteína conservada	0.988379	0.003623	0.00446	0.003538	SP(Sec/SPI)
O07415	Proteína de la familia Mce Mce1C	0.77221	0.017197	0.069873	0.14072	SP(Sec/SPI)
P9WLF1	Proteína no caracterizada Rv2277c	0.570253	0.011209	0.32456	0.093979	SP(Sec/SPI)
P9WL83	Proteína no caracterizada Rv2576c	0.728914	0.236321	0.004443	0.030322	SP(Sec/SPI)
I6Y204	D-alanil-D-alanina carboxipeptidasa DacB2 , D,D-carboxipeptidasa DacB2, DD-carboxipeptidasa, DD-peptidasa	0.993724	0.001116	0.004881	0.000279	SP(Sec/SPI)

P9WPQ5	Detiobiotina sintetasa BioD , DTB sintetasa, DTBS, Detiobiotina sintasa	0.456672	0.020149	0.133952	0.389227	SP(Sec/SPI)
P9WP81	Fosfoenolpiruvato transferasa , EPPG:FO PEP transferasa	0.813243	0.120471	0.009288	0.056998	SP(Sec/SPI)
P96271	Posible transporte de cationes metálicos ATPasa tipo P CtpH	0.522095	0.410326	0.020583	0.046996	SP(Sec/SPI)
I6Y4D2	N-acetilmuramoil-L-alanina amidasa Rv3717 , peptidoglicano amidasa dependiente de zinc	0.909068	0.02931	0.026514	0.035109	SP(Sec/SPI)
P9WN77	Glicerol-3-fosfato deshidrogenasa [NAD(P)+] , glicerol-3-fosfato deshidrogenasa dependiente de NAD(P)H	0.636659	0.045693	0.026778	0.290869	SP(Sec/SPI)
O06380	Carboxipeptidasa Rv3627c	0.582731	0.140566	0.146162	0.130542	SP(Sec/SPI)
P9WLR9	Proteína no caracterizada Rv1815	0.958342	0.039106	0.001908	0.000643	SP(Sec/SPI)
O07741	Proteína conservada	0.725655	0.257826	0.007091	0.009429	SP(Sec/SPI)
O53281	Probable acetiltransferasa Rv3034c , Proteína de homeostasis del peroxisoma Rv3034c	0.330464	0.310649	0.061499	0.297388	SP(Sec/SPI)
O69699	Proteína conservada	0.800749	0.025737	0.135469	0.038045	SP(Sec/SPI)
P9WNL9	Probable arabinosyltransferase A	0.470426	0.110333	0.101746	0.317495	SP(Sec/SPI)
O53168	Peptidoglicano endopeptidasa RipA , Invasión de macrófagos y proteína A de persistencia intracelular, Socio A de interacción del factor promotor de la reanimación, Proteína A que interactúa con Rpf	0.90401	0.061625	0.029095	0.005269	SP(Sec/SPI)
P9WGN3	Subunidad de proteína translocasa SecY	0.457074	0.080307	0.025973	0.436646	SP(Sec/SPI)
P9WKV9	Proteína no caracterizada Rv0477	0.995675	0.000383	0.003684	0.000258	SP(Sec/SPI)
Q79FR3	proteína de la familia PE PE13	0.457631	0.217698	0.027996	0.296675	SP(Sec/SPI)
Q6MWX4	probable proteína de membrana	0.748515	0.053655	0.17469	0.02314	SP(Sec/SPI)
P9WLQ1	Proteína no caracterizada Rv1987	0.777837	0.200062	0.00465	0.01745	SP(Sec/SPI)
P9WKD9	Proteína de unión final no homóloga Ku , Mt-Ku	0.872894	0.011783	0.009969	0.105355	SP(Sec/SPI)
O07787	Proteína de la familia Mce Mce2C	0.830278	0.062066	0.017567	0.090089	SP(Sec/SPI)

P9WP39	Probable carboxilesterasa Culp3 , proteína tipo cutinasa 3, Culp3	0.786277	0.154289	0.004665	0.054768	SP(Sec/SPI)
O05461	Mycosin-1 , proteasa MycP1	0.972331	0.001166	0.02607	0.000433	SP(Sec/SPI)
L0T243	Proteína de membrana probablemente conservada	0.419501	0.09467	0.177194	0.308634	SP(Sec/SPI)
P9WJT7	Probable proteína transportadora MmpL12	0.429288	0.201089	0.035791	0.333832	SP(Sec/SPI)
P9WJU1	Exportador de aciltrehalosa MmpL10	0.883729	0.014535	0.032983	0.068753	SP(Sec/SPI)
O53970	Proteína de la familia Mce Mce3D	0.895098	0.092064	0.003043	0.009795	SP(Sec/SPI)
P9WQJ1	Transportador ABC no caracterizado Proteína de unión a ATP Rv1273c	0.802748	0.022483	0.006835	0.167934	SP(Sec/SPI)
P9WN33	Putativa asparagina sintetasa [hidrolizante de glutamina]	0.430066	0.149286	0.030933	0.389716	SP(Sec/SPI)
P9WJT9	Proteína de captación de hemo MmpL11	0.471948	0.016584	0.086047	0.425421	SP(Sec/SPI)
P9WIQ3	NADPH-ferredoxina reductasa FprA , NFR	0.697198	0.021953	0.033274	0.247576	SP(Sec/SPI)
I6YGA5	Proteína hipotética asociada a Mce conservada	0.370974	0.008337	0.289213	0.331476	SP(Sec/SPI)
P9WKB9	Putativa diacilglicerol O-aciltransferasa Rv1760, putativa triacilglicerol sintasa Rv1760	0.634411	0.071653	0.036769	0.257167	SP(Sec/SPI)
I6X5C5	Posible oxidorreductasa rica en alanina	0.619386	0.116347	0.015616	0.248651	SP(Sec/SPI)
P71894	Lipoproteína de unión a azúcar periplásmica probable UspC	0.466400	0.062242	0.071186	0.400172	SP(Sec/SPI)
P9WPS5	Probable ATPasa I transportadora de cationes	0.540611	0.108945	0.032159	0.318285	SP(Sec/SPI)
O50383	Proteína que contiene el dominio DUF732	0.679364	0.318888	0.001320	0.000428	SP(Sec/SPI)
I6YB54	Proteína no caracterizada	0.693331	0.301081	0.003861	0.001727	SP(Sec/SPI)
P9WHU5	Peptidoglicano endopeptidasa RipB , Invasión de macrófagos y proteína B de persistencia intracelular, Socio B de interacción del factor promotor de la reanimación, Proteína B que interactúa con Rpf	0.981782	0.009460	0.006903	0.001855	SP(Sec/SPI)
P9WM43	Proteína no caracterizada Rv1271c	0.744637	0.197451	0.031703	0.026210	SP(Sec/SPI)

P9WI87	Pilin , subunidad estructural de Pili	0.986791	0.008464	0.004530	0.000216	SP(Sec/SPI)
P9WJS9	Proteína accesoria de exportación de sideróforos MmpS4 , PGB14T-X	0.512351	0.004323	0.084188	0.399137	SP(Sec/SPI)
P9WJG5	Proteína de membrana no caracterizada ArfC	0.515605	0.034605	0.089504	0.360286	SP(Sec/SPI)
P9WIW7	NADH-quinona oxidoreductasa subunidad A , NADH deshidrogenasa I subunidad A, NDH-1 subunidad A, NUO1	0.657622	0.041926	0.095200	0.205252	SP(Sec/SPI)
O05436	Proteína conservada	0.771240	0.220484	0.004224	0.004053	SP(Sec/SPI)
O53976	Proteína que contiene el dominio SCP	0.732053	0.264643	0.002545	0.000759	SP(Sec/SPI)
O53638	L,D-transpeptidasa 1 , LDT 1, Ldt(Mt1)	0.885668	0.004425	0.103444	0.006462	SP(Sec/SPI)
P9WKF3	Subunidad de unión de potasio de ATPasa transportadora de potasio , cadena A de ATP fosfohidrolasa [transportadora de potasio], subunidad A de unión y translocación de potasio, cadena A de ATPasa translocadora de potasio	0.777477	0.020497	0.035013	0.167013	SP(Sec/SPI)
O53538	Proteína rica en prolina hipotética conservada	0.471655	0.076392	0.092281	0.359672	SP(Sec/SPI)
O53634	Posible GDP-manosa 4,6-deshidratasa Gca (GDP-D-manosa deshidratasa)	0.698247	0.015969	0.032268	0.253516	SP(Sec/SPI)
P9WKR7	Proteína no caracterizada Rv0875c	0.829834	0.005667	0.105347	0.059152	SP(Sec/SPI)
Q6MWY0	Proteína de la familia PE-PGRS PE_PGRS50	0.43004	0.164843	0.035886	0.369231	SP(Sec/SPI)
O86365	Glucosidasa no caracterizada Rv0584 ,	0.856627	0.034193	0.019829	0.089351	SP(Sec/SPI)
O53383	Proteína rica en prolina hipotética	0.890796	0.069171	0.010983	0.029051	SP(Sec/SPI)
P96408	Proteína transmembrana probablemente conservada	0.523877	0.040987	0.082153	0.352983	SP(Sec/SPI)
O53380	D-alanil-D-alanina carboxipeptidasa DacB1 , D,D-carboxipeptidasa DacB1, DD-carboxipeptidasa, DD-peptidasa	0.82184	0.1332	0.041172	0.003788	SP(Sec/SPI)
O69694	Posible oxidoreductasa	0.636689	0.129984	0.024306	0.209021	SP(Sec/SPI)

P96419	Alfa-(1->3)- arabinofuranosiltransferasa, Arabinofuranan 3-O- arabinosiltransferasa	0.815735	0.063708	0.085284	0.035274	SP(Sec/SPI)
P9WHU3	Probable endopeptidase Rv2190c	0.904553	0.064685	0.005948	0.024814	SP(Sec/SPI)
P9WNG1	Monooxigenasa no caracterizada Rv0892	0.590707	0.014963	0.203574	0.190755	SP(Sec/SPI)
I6WZH9	Proteína exportada probablemente conservada	0.614592	0.012506	0.05215	0.320753	SP(Sec/SPI)
L0TFC2	Proteína de la familia PE-PGRS PE_PGRS49	0.535779	0.078052	0.048868	0.337302	SP(Sec/SPI)
O06823	Prolina probablemente conservada, glicina, proteína secretada rica en valina	0.851944	0.140313	0.002245	0.005499	SP(Sec/SPI)
O86370	Posible proteína exportada conservada	0.877977	0.030221	0.029946	0.061856	SP(Sec/SPI)
O06394	Proteína transmembrana probablemente conservada	0.717957	0.058773	0.098069	0.125201	SP(Sec/SPI)
I6YC99	Proteína de la familia Mce Mce4A	0.617433	0.178241	0.035242	0.169084	SP(Sec/SPI)
O86313	Posible proteína de membrana	0.736413	0.163186	0.032577	0.067824	SP(Sec/SPI)
P9WfZ9	Putativa proteína permeasa transportadora de péptidos Rv1282c	0.641138	0.208429	0.014069	0.136364	SP(Sec/SPI)
O06555	Proteína rica en ala conservada	0.6559	0.315419	0.015016	0.013665	SP(Sec/SPI)
P9WHZ5	Proteína de la familia PPE no caracterizada PPE42	0.398199	0.30821	0.025819	0.267773	SP(Sec/SPI)
O53701	Proteína que contiene el dominio Zeta_toxin	0.816654	0.024381	0.004657	0.154308	SP(Sec/SPI)
O07423	Posible lipoproteína LprO	0.445948	0.222796	0.324799	0.006457	SP(Sec/SPI)
P9WIM5	Ramnosil O-metiltransferasa	0.634113	0.174421	0.011617	0.179849	SP(Sec/SPI)
Q79FV3	Proteína de la familia PE-PGRS PE_PGRS15	0.416911	0.214821	0.017762	0.350505	SP(Sec/SPI)
P9WFP1	UPF0053 proteína Rv2366c	0.615202	0.04668	0.074693	0.263426	SP(Sec/SPI)
L0T3X8	Proteína de la familia PE-PGRS PE_PGRS6 , PE_PGRS6	0.346227	0.333795	0.022946	0.297032	SP(Sec/SPI)
P9WKQ1	Esfingomielinasa , SMasa	0.919407	0.051126	0.023597	0.00587	SP(Sec/SPI)
P9WHR5	Carboxilesterasa B	0.651012	0.333717	0.013503	0.001768	SP(Sec/SPI)
I6WZ26	Proteína no caracterizada	0.87517	0.040328	0.007728	0.076774	SP(Sec/SPI)

P9WKD1	Peptidoglicano D,D-transpeptidasa PbpA , proteína A de unión a penicilina, PBPA	0.891379	0.008438	0.047784	0.052399	SP(Sec/SPI)
O50440	Diaciltrehalosa aciltransferasa Chp2	0.971988	0.004667	0.019862	0.003484	SP(Sec/SPI)
I6X5W6	Posible membrana conservada o proteína exportada	0.619146	0.274583	0.075978	0.030293	SP(Sec/SPI)
O69659	Proteína de membrana no caracterizada Rv3691	0.59443	0.136332	0.064688	0.20455	SP(Sec/SPI)
P9WJT5	Probable proteína accesoria de transporte MmpS1	0.848069	0.012411	0.040054	0.099467	SP(Sec/SPI)
L7N695	inmunomodulador de la familia PE PE5	0.484938	0.299586	0.027611	0.187865	SP(Sec/SPI)
P9WFZ7	Putativata proteína permeasa transportadora de péptidos Rv1283c	0.460362	0.091832	0.288865	0.158941	SP(Sec/SPI)
I6XHD1	Proteína no caracterizada	0.974912	0.019198	0.004002	0.001887	SP(Sec/SPI)
I6Y4F1	Proteína transmembrana probablemente conservada	0.985889	0.013174	0.000617	0.000319	SP(Sec/SPI)
O05444	Posible proteína de membrana	0.476653	0.076292	0.04129	0.405765	SP(Sec/SPI)
O07775	Posible proteína exportada	0.870074	0.082724	0.023465	0.023737	SP(Sec/SPI)
P9WG31	Factor promotor de la reanimación RpfA	0.702491	0.291846	0.003278	0.002384	SP(Sec/SPI)
O53179	Proteína no caracterizada	0.532564	0.240419	0.110275	0.116743	SP(Sec/SPI)
O06787	probable proteína de membrana	0.418661	0.321329	0.030495	0.229514	SP(Sec/SPI)
P9WLL5	Proteína no caracterizada Rv2075c	0.911211	0.085325	0.002575	0.000889	SP(Sec/SPI)
Q79FT3	Proteína de la familia PE-PGRS PE_PGRS19	0.461208	0.286172	0.026059	0.226561	SP(Sec/SPI)
P96226	Posible proteína de membrana	0.76032	0.1873	0.012093	0.040286	SP(Sec/SPI)
Q6MWW7	Proteína de la familia PE-PGRS PE_PGRS56	0.554091	0.066039	0.024981	0.354889	SP(Sec/SPI)
P95210	Proteína transmembrana probablemente conservada	0.308862	0.28434	0.111161	0.295636	SP(Sec/SPI)
L7N652	Probable proteína de membrana integral transportadora de azúcar ABC transportador UspB	0.493343	0.140499	0.028833	0.337325	SP(Sec/SPI)
P9WG27	Factor promotor de la reanimación RpfD	0.680801	0.054874	0.013409	0.250916	SP(Sec/SPI)
P9WL33	Proteína no caracterizada Rv2891	0.967004	0.029919	0.001884	0.001194	SP(Sec/SPI)

L7N667	proteína de la familia PE PE8	0.395814	0.383768	0.017327	0.203092	SP(Sec/SPI)
P95133	Proteína no caracterizada	0.4752	0.297038	0.020167	0.207595	SP(Sec/SPI)
O06545	Probable proteína de transporte transmembrana conservada MmpL13a	0.698334	0.035203	0.029026	0.237438	SP(Sec/SPI)
Q79FP3	Proteína de la familia PE-PGRS PE_PGRS26	0.581307	0.156246	0.040227	0.22222	SP(Sec/SPI)
P9WJY7	Probable transportador de nitrato/nitrito NarK2	0.511667	0.022881	0.010843	0.45461	SP(Sec/SPI)
O53177	Factor promotor de la reanimación RpfE	0.966306	0.029485	0.002612	0.001597	SP(Sec/SPI)
O53975	Proteína de membrana probablemente conservada	0.852777	0.130295	0.014051	0.002878	SP(Sec/SPI)
P71912	Proteína transmembrana conservada	0.620793	0.195293	0.007768	0.176146	SP(Sec/SPI)
P9WM47	Proteína no caracterizada Rv1268c	0.892311	0.105161	0.001275	0.001253	SP(Sec/SPI)
O69623	Inhibidor de la apoptosis Rv3655c	0.670917	0.019404	0.127713	0.181966	SP(Sec/SPI)
O06825	Probable L,D-transpeptidasa 3 , LDT 3, Ldt(Mt3)	0.882538	0.007993	0.040942	0.068527	SP(Sec/SPI)
Q6MWX3	Posible dTDP-glucosa 4,6-deshidratasa	0.457534	0.28901	0.007877	0.245579	SP(Sec/SPI)
O53148	Probable transportador ABC de membrana integral de transporte de antibióticos no identificado	0.463897	0.131311	0.060243	0.344549	SP(Sec/SPI)
I6YC58	Mycosin-4 , MycP4 proteasa	0.854512	0.125612	0.018915	0.000961	SP(Sec/SPI)
I6Y3N9	Proteína conservada	0.993872	0.005119	0.00072	0.00029	SP(Sec/SPI)
P9WM45	Proteína Rv1269c	0.196112	0.802813	0.000747	0.000327	TAT(Tat/SPI)
P9WK21	Metionina aminopeptidasa 1 , MAP 1, MetAP 1, Peptidasa M	0.145003	0.525417	0.011625	0.317955	TAT(Tat/SPI)
O07242	Posible beta-1,3-glucanasa	0.006412	0.991546	0.001833	0.000209	TAT(Tat/SPI)
P9WQP1	Diacilglicerol aciltransferasa/micoliltransferasa Ag85B , DGAT, proteína extracelular de 30 kDa, Acil-CoA:diacilglicerol aciltransferasa, antígeno 85 complejo B, 85B, Ag85B, antígeno alfa extracelular, proteína B de unión a fibronectina, Fbps B	0.422641	0.553297	0.015434	0.008628	TAT(Tat/SPI)

P9WQ79	4-aminobutirato aminotransferasa, (S)-3-amino-2-metilpropionato transaminasa, GABA aminotransferasa, GABA-AT, gamma-amino-N-butilato transaminasa, GABA transaminasa, glutamato: transaminasa semialdehído succínico, L-AIBAT	0.144507	0.51182	0.004233	0.33944	TAT(Tat/SPI)
P9WQP3	Diacilglicerol aciltransferasa/micoliltransferasa Ag85A, DGAT, Acil-CoA:diacilglicerol aciltransferasa, antígeno 85 complejo A, 85A, Ag85A, proteína A de unión a fibronectina, Fbps A	0.137075	0.837983	0.008355	0.016587	TAT(Tat/SPI)
L7N6B2	Probable lipoproteína periplásmica transportadora de hierro	0.001101	0.526419	0.471883	0.000596	TAT(Tat/SPI)
O53608	Posible oxidorreductasa	0.028723	0.969504	0.001073	0.0007	TAT(Tat/SPI)
P96272	Posible proteína transmembrana	0.359173	0.425369	0.032481	0.182976	TAT(Tat/SPI)
I6WZK7	Oxidasa multicobre MmcO, Oxidasa multicobre de Mycobacterium	0.005116	0.960144	0.034184	0.000556	TAT(Tat/SPI)
P9WHK7	CTP sintasa, citidina 5'-trifosfato sintasa, citidina trifosfato sintetasa, CTP sintetasa, CTPS, UTP-amoniaco ligasa	0.252539	0.381119	0.044827	0.321515	TAT(Tat/SPI)
P9WIY3	8-oxo-(d)GTP fosfatasa, 8-oxo-(d)GTPasa, MtuMutT1, proteína mutante MutT1	0.314998	0.419672	0.013901	0.251429	TAT(Tat/SPI)
O06619	Proteína conservada	0.048815	0.507492	0.010072	0.433621	TAT(Tat/SPI)
P9WFI3	Putativa metiltransferasa Rv0830 dependiente de S-adenosil-L-metionina	0.239403	0.694886	0.002003	0.063708	TAT(Tat/SPI)
P9WPG9	Probable CDP-diacilglicerol pirofosfatasa, CDP-diacilglicerol fosfatidilhidrolasa, CDP-diglicérido hidrolasa	0.422428	0.522727	0.033508	0.021337	TAT(Tat/SPI)
P72062	Proteína conservada	0.019033	0.558594	0.411696	0.010677	TAT(Tat/SPI)
P9WP41	Probable carboxilesterasa Culp2, CFP25, proteína tipo cutinasa 2, Culp2	0.146588	0.836631	0.007254	0.009527	TAT(Tat/SPI)

P9WKP3	Proteína no caracterizada Rv0906	0.428987	0.526896	0.014596	0.029521	TAT(Tat/SPI)
P9WLS1	Proteína no caracterizada Rv1813c	0.304557	0.675062	0.004216	0.016166	TAT(Tat/SPI)
P96809	Deshidrogenasa de ácido hidroximicólico dependiente de F420 , fHMAD, 1 FGD2	0.00057	0.8366	0.162463	0.000367	TAT(Tat/SPI)
P9WIB3	Fosfolipasa CB , PLC-B, Fosfolipasa CB micobacteriana	0.002813	0.996506	0.000245	0.000437	TAT(Tat/SPI)
P9WG63	Proteína DipZ	0.021838	0.844054	0.049463	0.084645	TAT(Tat/SPI)
I6XEI5	Hidrolasa de peptidoglicano putativa Rv2525c	0.04897	0.94117	0.005307	0.004553	TAT(Tat/SPI)
O53410	Proteína que contiene el dominio PNPLA	0.11538	0.781601	0.01358	0.089438	TAT(Tat/SPI)
I6WZD9	Posible oxidorreductasa	0.060053	0.759436	0.013656	0.166855	TAT(Tat/SPI)
P9WKV3	L,D-transpeptidasa 5 , LDT 5, Ldt(Mt5)	0.096417	0.444849	0.087259	0.371475	TAT(Tat/SPI)
P9WIB5	Fosfolipasa CA , PLC-A, antígeno MTP40, fosfolipasa CA micobacteriana	0.00855	0.990198	0.000326	0.000926	TAT(Tat/SPI)
P9WIB1	Fosfolipasa CC , PLC-C	0.005839	0.99326	0.000383	0.000519	TAT(Tat/SPI)
P9WL69	Proteína no caracterizada Rv2599	0.360646	0.375143	0.107007	0.157203	TAT(Tat/SPI)
L7N648	proteína de la familia PE PE6	0.219256	0.402739	0.048874	0.32913	TAT(Tat/SPI)
O05442	Proteína del canal de la membrana externa CpnT , proteína del canal con toxina inductora de necrosis	0.148216	0.392166	0.311401	0.148217	TAT(Tat/SPI)
P9WLD9	Proteína no caracterizada Rv2297	0.423526	0.455051	0.015957	0.105466	TAT(Tat/SPI)
P9WHW9	inmunomodulador de la familia PPE PPE68	0.357213	0.486529	0.021528	0.13473	TAT(Tat/SPI)
P71619	Probable lipoproteína de unión a Sn-glicerol-3-fosfato UgpB	0.000542	0.912494	0.086589	0.000375	TAT(Tat/SPI)
P96901	Probable helicasa dependiente de ATP Lhr (proteína relacionada con helicasa grande)	0.326481	0.442386	0.010403	0.220731	TAT(Tat/SPI)
O50419	Proteína que contiene el dominio UmuC	0.159275	0.595883	0.074982	0.169860	TAT(Tat/SPI)
I6Y8R4	Acil-CoA:diacilglicerol aciltransferasa	0.056717	0.922972	0.018581	0.001730	TAT(Tat/SPI)
I6YAE2	Probable proteína rica en alanina transmembrana conservada	0.009850	0.622528	0.365290	0.002332	TAT(Tat/SPI)

P9WQM3	Transportador no caracterizado Rv1999c	0.086801	0.501238	0.009547	0.402414	TAT(Tat/SPI)
O07246	Posible proteína exportada conservada	0.461468	0.511804	0.013403	0.013325	TAT(Tat/SPI)
O53200	Posible proteína de membrana rica en alanina y prolina	0.17457	0.457057	0.040964	0.32741	TAT(Tat/SPI)
P9WHZ7	Proteína de la familia PPE no caracterizada PPE40	0.145088	0.429289	0.008737	0.416886	TAT(Tat/SPI)
Q6MWX5	Proteína de la familia PE-PGRS PE_PGRS52	0.313113	0.437501	0.016825	0.232561	TAT(Tat/SPI)
O53485	Probable lipoproteína fijadora de azúcar	0.009384	0.850826	0.138748	0.001042	TAT(Tat/SPI)
P9WFJ1	Putativa metiltransferasa Rv0145 dependiente de S-adenosil-L-metionina	0.177418	0.423736	0.014705	0.384141	TAT(Tat/SPI)
P9WGR7	Oxidorreductasa no caracterizada Rv0945	0.041987	0.85484	0.004447	0.098727	TAT(Tat/SPI)
P95046	Proteína no caracterizada	0.091143	0.584611	0.009734	0.314512	TAT(Tat/SPI)
O33364	Acil-CoA:diacilglicerol aciltransferasa	0.016794	0.975948	0.001312	0.005946	TAT(Tat/SPI)
O53836	Proteína conservada	0.066893	0.693876	0.004563	0.234668	TAT(Tat/SPI)
O53734	Proteína que contiene el dominio amino_oxidasa	0.146125	0.365094	0.16892	0.319861	TAT(Tat/SPI)
L0T7Y7	Proteína de la familia PPE PPE31	0.165555	0.443834	0.010923	0.379688	TAT(Tat/SPI)
I6X6H8	Proteína de la familia PPE PPE52	0.229058	0.461407	0.022622	0.286913	TAT(Tat/SPI)
O05888	Proteína no caracterizada	0.397031	0.502415	0.006656	0.093897	TAT(Tat/SPI)
P9WP63	Transportador de iones de fluoruro putativo CrcB 1	0.093507	0.664637	0.014734	0.227122	TAT(Tat/SPI)
P9WL81	Proteína no caracterizada Rv2577	0.000619	0.997996	0.000422	0.000963	TAT(Tat/SPI)
Q6MWW0	Proteína de la familia PPE PPE64	0.212988	0.425352	0.01261	0.34905	TAT(Tat/SPI)
O50382	Proteína no caracterizada	0.195926	0.770273	0.005386	0.028415	TAT(Tat/SPI)
P9WM77	Proteína no caracterizada Rv0057	0.077899	0.57113	0.026798	0.324173	TAT(Tat/SPI)
O53423	Esterasa similar a GDSL Rv1075c , Acetilesterasa	0.136558	0.789953	0.040092	0.033396	TAT(Tat/SPI)
P9WHX7	Proteína de la familia PPE no caracterizada PPE62	0.311044	0.420002	0.017557	0.251398	TAT(Tat/SPI)
P9WI31	Proteína de la familia PPE PPE15 , perilipina-1 micobacteriana, MPER1	0.212201	0.386718	0.015591	0.385491	TAT(Tat/SPI)

*Predicción en base al mayor valor de probabilidad de presentar o no un péptido señal:

- Sec/SPI: péptidos señal secretores "estándar" transportados por el translocón Sec y escindidos por la péptido señal I (Lep)
- Sec/SPII: péptidos señal de lipoproteínas transportados por el translocón Sec y escindidos por Signal Peptidase II (Lsp)
- Tat/SPI: péptidos señal Tat transportados por el translocón Tat y escindidos por la péptido señal I (Lep)
- OTHER: si este valor es alto, la proteína no presenta péptido señal.

Anexo 4. Proteínas del proteoma predichas de secreción por la vía no clásica (ausencia de Péptido señal) mediante el programa SecretomeP (n=630).

Uniprot ID	Nombre de la proteína	SecP score*
P96376	Posible proteína de membrana conservada	0.501183
Q79FZ9	Proteína de la familia Mce Mce1A	0.503792
I6X7F9	Regulador transcripcional Rv3488	0.504440
Q79G13	Glucanasa	0.504851
P95225	Antitoxina putativa Rv0268c	0.506329
P9WP17	dCTP desaminasa, formadora de dUMP , dCTP desaminasa bifuncional:dUTPasa, DCD-DUT	0.508095
O05310	Proteína conservada	0.508661
P9WL93	Proteína no caracterizada Rv2569c	0.511788
P71914	Ciclase probable (Adenilil-o guanilil-)(Adenilato-o guanilato-)	0.516681
Q79FG2	Proteína rica en glicina hipotética	0.520330
L7N694	Proteína no caracterizada	0.520429
P9WH63	30S proteína ribosomal S12	0.520491
P9WNP9	Sistema de secreción ESX-5 ATPasa EccB5, componente B5 conservado de ESX, proteína del sistema de secreción tipo VII EccB5, proteína EccB5 de T7SS	0.521825
I6X9F4	Posible proteína de membrana	0.524227
I6XDU8	Proteína conservada	0.528004
O53969	Proteína de la familia Mce Mce3C	0.529428
O06145	Proteína conservada	0.530354
P9WNB3	ESX-1 proteína del sistema de secreción EccCa1 , componente conservado ESX Ca1, proteína secretora Snm1, proteína del sistema de secreción Tipo VII EccCa1, proteína T7SS EccCa1	0.531095
I6YGC8	Probable ácido graso-CoA ligasa FadD18 (ácido graso-CoA sintetasa) (ácido graso-CoA sintetasa)	0.534102

P9WN39	Glutamina sintetasa , GS, Glutamato-amoniaco ligasa, Glutamina sintetasa I beta, GSI beta	0.539778
P9WGD5	Proteína de unión al ADN monocatenario, SSB	0.541185
P9WL11	Proteína no caracterizada Rv3073c	0.541245
I6YB06	Proteína similar a la glutaredoxina NrdH	0.543958
I6WZI4	Proteína conservada	0.546645
P9WJM5	Péptido metionina sulfóxido reductasa MsrA , Proteína-metionina-S-óxido reductasa, Péptido-metionina (S)-S-óxido reductasa, Péptido Met(O) reductasa	0.547274
Q6MX36	Proteína no caracterizada	0.547687
P9WH53	30S proteína ribosomal S16	0.550636
P9WPV5	Subunidad b de la ATP sintasa , subunidad b del sector F(0) de la ATP sintasa, subunidad I de la ATPasa, subunidad b de la ATPasa tipo F, subunidad b de la F-ATPasa	0.559020
I6Y8I5	Enzima generadora de formilglicina, FGE	0.562146
P9WHW5	Serina/treonina proteína fosfatasa PstP, fosfatasa Ser/Thr micobacteriana, Mstp, fosfatasa Ser/Thr de la familia PP2C	0.564715
O69624	Proteína no caracterizada	0.567870
O06537	Proteína no caracterizada	0.570614
Q79FA4	Probable carboxilesterasa Culp7 , proteína tipo cutinasa 7, Culp7	0.571772
P9WNR3	Sistema de secreción ESX-3 ATPasa EccB3, componente B3 conservado de ESX, proteína EccB3 del sistema de secreción tipo VII, proteína EccB3 de T7SS	0.581724
O05447	Proteína no caracterizada	0.582019
P71846	Posible proteína de transferencia de electrones FdxB	0.582665
P9WHG5	Toxina ParE2	0.585496
P9WHR9	Serina proteasa Rv3671c	0.592974
P9WLK5	Proteína no caracterizada Rv2081c	0.593091
Q79FD9	Proteína no caracterizada	0.594228
P9WJF3	Proteína de síntesis de la pared celular CwsA, Síntesis de la pared celular y proteína A de forma celular	0.595243
O53178	Proteína hipotética rica en prolina y serina	0.595772
P9WIU5	Proteína de unión a peptidoglicano ArfA, Porina A de la membrana externa, Proteína A de la membrana externa, OmpATb, Proteína ArfA de la membrana externa	0.595927
P9WI77	Serina/treonina-proteína quinasa PknE	0.599320
P9WQ45	Ácido graso putativo--CoA ligasa fadD25 , acil-CoA sintetasa	0.600086
P9WLK3	Proteína no caracterizada Rv2083	0.600204

I6X7G4	Proteína conservada hipotética rica en alanina y valina asociada a Mce	0.605931
P9WLF7	Proteína no caracterizada Rv2271	0.609158
P9WNM3	Factor de elongación P, EF-P	0.609414
I6X7H4	Ferredoxina	0.613163
P9WJD9	Proteína asociada a la secreción de ESX-1 EspB, Antígeno MTB48	0.613343
O53937	Probable ferredoxina	0.613644
P0DN33	Proteína no caracterizada Rv0609B	0.615743
P9WP15	Nitroreductasa dependiente de deazaflavina , quinona reductasa dependiente de F420H(2) Ddn, Fqr	0.617107
O33270	Proteína conservada	0.619394
P9WPC1	Proteína adaptadora de proteasa Clp dependiente de ATP ClpS	0.621210
P9WH41	30S proteína ribosómica S20	0.624532
O53973	Proteína de membrana asociada a Mce probablemente conservada	0.626171
P9WPS7	Probable ATPasa G transportadora de cationes	0.627981
P9WFU7	Biosíntesis de lisilfosfatidilglicerol proteína bifuncional LysX, lisina--tRNA ligasa, lisil-tRNA sintetasa, LysRS, fosfatidilglicerol lisiltransferasa, 2.3.2.3, lisilfosfatidilglicerol sintetasa, LPG sintetasa	0.628024
P9WHX1	Proteína de la familia PPE no caracterizada PPE66	0.628026
P9WI63	Serina/treonina-proteína quinasa PknL	0.630743
L0T9Q6	Proteína no caracterizada	0.630822
O06136	Proteína de membrana conservada	0.633053
P9WL57	Proteína no caracterizada Rv2635	0.639470
P9WIA9	CD de fosfolipasa , PLC-D	0.641782
P96912	Posible proteína de membrana	0.642954
O07777	Sensor histidina quinasa componente HK2	0.644335
I6Y1I5	Posible proteína rica en alanina secretada conservada	0.644372
O06138	Probable 'componente vinculado con el ensamblaje del citocromo' transporte transmembrana proteína de unión a ATP transportador ABC CydD	0.646634
I6Y8V2	Probable tiorredoxina ThiX	0.646974
P9WNK7	Diana antigénica secretora temprana de 6 kDa, ESAT-6	0.647347
I6YGW6	Posible proteína similar a la tiorredoxina anclada a la membrana (proteína relacionada con el intercambio de tiol-disulfuro)	0.647440

O53402	integrasa probable	0.647470
I6X9Y7	Proteína conservada	0.654371
P9WGY5	Subunidad omega de ARN polimerasa dirigida por ADN, subunidad omega RNAP , subunidad omega de ARN polimerasa, subunidad omega de transcriptasa	0.654418
Q50681	Probable éster carboxílico hidrolasa LipM	0.656502
Q6MWX8	Proteína de la familia PPE PPE56	0.657643
O53401	Proteína no caracterizada	0.658359
P9WI69	Serina/treonina-proteína quinasa PknI	0.662510
P9WHA1	50S proteína ribosomal L31	0.663659
P9WFJ7	UPF0353 proteína Rv1481	0.663785
O33272	Posible membrana conservada o proteína secretada	0.664017
P71820	Posible ferredoxina	0.665143
Q6MWX9	Proteína de la familia PPE PPE55	0.665482
P71734	Proteína conservada	0.666761
P9WK09	Metalotioneína , MT, metalotioneína de unión a cobre, metalotioneína de unión a Cu(I), metalotioneína micobacteriana	0.667173
P71915	Proteína transmembrana probablemente conservada	0.667488
O69731	Proteína no caracterizada Rv3766	0.669922
I6XG13	Proteína reguladora de la transcripción probable (probablemente de la familia AsnC)	0.670015
O07784	Proteína de la familia Mce Mce2F	0.674457
P9WG17	Permeasa transportadora ABC no caracterizada Rv0072	0.676042
P9WPPE5	Co-chaperonina GroES, antígeno de 10 kDa, chaperonina de 10 kDa, proteína de choque térmico BCG-A, chaperonina-10, Cpn10	0.676991
O05439	Proteína no caracterizada	0.677167
L7N664	tiorredoxina	0.678157
P9WI49	Proteína de la familia PPE no caracterizada PPE1	0.678821
P9WJQ1	Proteína similar a la nitrato reductasa NarX	0.678967
P9WJC1	Proteína asociada a la secreción de ESX-1 EspK	0.682592
P9WMT5	Proteína GrpE, cofactor HSP-70	0.684659
P9WJA9	Regulador de acumulación de glucógeno GarA	0.687728
O06291	Probable serina proteasa HtrA1 , Requisito de alta temperatura A proteasa	0.688235
P9WI81	Serina/treonina-proteína quinasa PknB	0.691479

P9WNA3	ADN translocasa FtsK	0.692323
P9WM87	Proteína no caracterizada Rv0048c	0.692865
P96233	Proteína no caracterizada	0.693382
P9WM71	Proteína no caracterizada Rv0090	0.694818
P96258	Posible proteína de membrana conservada	0.695169
O53581	Carboxilesterasa/lipasa Culp6 , lipasa de la pared celular, proteína tipo cutinasa 6, Culp6	0.695352
O69625	Posible proteína de membrana rica en alanina conservada	0.695499
P9WPT7	Probable ATPasa J tipo P transportadora de cationes	0.695890
O53840	Posible transposasa	0.697761
P9WJB5	Proteína que contiene el dominio FHA FhaB , proteína A que interactúa con FtsZ	0.699358
P96277	Péptido putativo relacionado con la tuberculina	0.700751
P9WG15	Permeasa transportadora ABC no caracterizada Rv2563	0.702524
P71810	Exportación probable o proteína de membrana	0.704631
P9WHF9	Putativa tiosulfato sulfurtransferasa , proteína tipo rodanasa	0.705720
I6YC95	Proteína de la familia Mce Mce4F	0.706595
O05769	Proteína que contiene el dominio PHB	0.708338
O53493	Apolipoproteína N-aciltransferasa bifuncional/poliprenol monofosfomanosa sintasa , Apolipoproteína N-aciltransferasa, ALP N-aciltransferasa, 2.3.1.269, Poliprenol monofosfomanosa sintasa, PPM sintasa, Poliprenol-P-Man sintasa, Ppm1, 2.4.1.-, Dolicol-fosfato manosa sintasa, 2.4.1.83	0.709089
O06775	Putativa antitoxina VapB8	0.712269
P9WNR7	Sistema de secreción ESX-1 ATPasa EccB1 , componente B1 conservado de ESX, proteína EccB1 del sistema de secreción tipo VII, proteína EccB1 de T7SS	0.713526
P9WI35	Proteína de la familia PPE no caracterizada PPE13	0.714397
O06796	Posible proteína exportada	0.714853
P9WI79	Serina/treonina-proteína quinasa PknD	0.716154
I6Y3V3	Posible proteína transmembrana conservada	0.716675
P9WFN1	UPF0098 proteína Rv2140c	0.717439
P9WJD3	Proteína asociada a la secreción de ESX-1 EspE	0.717682
P95205	Proteína de la familia de repeticiones 13E12 conservada	0.718944
P9WFL5	UPF0182 proteína Rv0064	0.719590

P9WHS3	Metaloproteasa de zinc Rip1 , Regulador de proteasas sigma KLM, endopeptidasa S2P, proteasa Site-2 Rip1, proteasa S2P Rip1, proteasa intramembrana tipo Site-2	0.721527
P71799	Proteína de membrana probablemente conservada	0.722589
P9WHJ5	Proteína de unión a ARN polimerasa RbpA	0.723416
Q79F96	Proteína que contiene el dominio PGRP	0.732491
P72005	Proteína no caracterizada	0.733059
O07239	Proteína rica en prolina y treonina hipotética conservada	0.734064
O07416	Proteína de la familia Mce Mce1D	0.734067
P9WPX7	3-deshidroquinato deshidratasa , 3-deshidroquinasa, tipo II DHQase	0.737256
O33249	Proteína conservada	0.737360
O53718	Posible proteína secretada conservada	0.738641
P71728	Proteína no caracterizada	0.738722
P9WM29	Proteína no caracterizada Rv1312	0.740738
P9WI41	Proteína de la familia PPE PPE10	0.742658
O53651	Posible proteína transmembrana conservada	0.743230
P9WI43	proteína de la familia PPE PPE4	0.743757
P9WPT9	ATPasa B tipo P transportadora de cationes	0.745186
O69658	Proteína de membrana probablemente conservada	0.745607
P9WPS3	Probable ATPasa V tipo P exportadora de cobre , ATPasa exportadora de Cu(+)	0.746696
I6X562	Proteína no caracterizada	0.747448
O50455	Proteína de transporte de magnesio CorA	0.747452
P9WPT3	Probable ATPasa tipo P exportadora de cobalto/níquel , ATPasa tipo P transportadora de cationes CtpD	0.747880
O33268	Probable reductasa de unión a hierro y azufre	0.748171
I6YA50	Putativa proteína del sistema de conservación de la envoltura Rv2743c	0.750078
O05781	Posible proteína transmembrana conservada	0.753475
P9WP13	Quinona reductasa dependiente de F420H(2) Rv1261c , Fqr	0.753821
P71833	Posible proteína transmembrana conservada	0.764187
Q79FK9	Proteína de la familia PE-PGRS PE_PGRS31	0.764378
P9WGN7	Subunidad de proteína translocasa SecE	0.764578
Q6MX48	proteína de la familia PPE PPE6	0.766551
P96923	Posible proteína exportada	0.769765

P9WM35	Proteína no caracterizada Rv1290c	0.769925
P9WJC5	Proteína asociada a la secreción de ESX-1 EspI	0.770243
P9WMX5	Glicosiltransferasas putativas	0.773771
P9WFL3	UPF0182 proteína Rv3193c	0.774026
P9WLS9	Proteína de membrana probable Rv1733c	0.776124
P9WJU5	Exportador de sulfolípidos-1 MmpL8	0.779982
P95164	Posible proteína sensora-transdutora BlaR	0.780062
I6XFZ8	Proteína Rv3035	0.782143
O53900	Proteína permeasa transportadora ABC no caracterizada Rv0987	0.783029
P71767	Proteína de membrana conservada	0.784180
O69695	Proteína de membrana de dos dominios probablemente conservada	0.784940
I6XVZ6	probable proteína de membrana	0.785459
I6WYT7	Proteína de la familia Mce Mce2D	0.787487
P0DMM4	Proteína no caracterizada Rv0572A	0.787938
I6Y479	Proteína transmembrana probablemente conservada	0.789167
P9WQI9	Los compuestos hidrofílicos importan la proteína de permeasa/unión de ATP BacA, Compuestos hidrófilos transportador ABC BacA	0.790893
I6Y2I9	Proteína de membrana integral probablemente conservada	0.792463
P9WIF7	Proteína de la familia PE-PGRS no caracterizada PE_PGRS24	0.792905
I6XI16	Proteína conservada	0.793671
P9WLQ9	Proteína no caracterizada Rv1836c	0.793740
O33222	probable proteína transmembrana	0.794011
O53930	Posible proteína de membrana integral	0.795303
P9WLY1	Proteína no caracterizada Rv1417	0.795452
Q50655	Proteína que contiene el dominio HNHc	0.795644
P9WJ07	Antitoxina Rv0909	0.796299
I6YA32	Hidrolasa de pared celular ChiZ	0.798463
O07420	Proteína transmembrana asociada a Mce probablemente conservada	0.800316
P9WLI1	Proteína no caracterizada Rv2219	0.800833
P9WIG5	antígeno WAG22	0.800969
O05857	Proteína rica en prolina y treonina hipotética conservada	0.801553

P9WJE3	Proteína EccE5 del sistema de secreción ESX-5 , componente E5 conservado de ESX , proteína EccE5 del sistema de secreción tipo VII, proteína EccE5 de T7SS	0.802584
P9WG99	Proteína translocasa proteína independiente de sec TatB	0.802863
O53621	probable proteína de membrana	0.806459
O06422	Proteína transmembrana probablemente conservada	0.809263
P9WJU9	Probable proteína transportadora MmpL6	0.810134
Q79FU2	Proteína de la familia PE-PGRS PE_PGRS17	0.811467
O53650	Proteína de membrana probablemente conservada	0.811785
O05900	Proteína conservada	0.811983
P9WHW3	Peptidil-prolil cis-trans isomerasa A, PPIasa A, ciclofilina, rotamasa A	0.812297
P9WPU3	ATPasa transportadora de potasio Subunidad de unión a ATP , Cadena B de ATP fosfohidrolasa [transportadora de potasio], Subunidad B de unión y translocación de potasio, Cadena B de ATPasa translocadora de potasio	0.812587
P9WKA1	Señal peptidasa I , SPasa I, líder peptidasa I	0.812770
P9WIE1	Peroxirredoxina putativa Rv2521 , proteína comigratoria de bacterioferritina, tiorredoxina peroxidasa, peroxirredoxina dependiente de tiorredoxina Rv2521	0.813804
P9WJE5	Proteína EccE3 del sistema de secreción ESX-3 , componente E3 conservado de ESX , proteína EccE3 del sistema de secreción tipo VII, proteína EccE3 de T7SS	0.816010
P9WKW5	Proteína de membrana no caracterizada Rv3835	0.816395
Q79FE4	Posible proteína de membrana conservada	0.816684
P9WH23	Citocromo bc1 complejo Rieske subunidad de hierro-azufre , citocromo bc1 reductasa complejo subunidad QcrA, proteína de hierro-azufre de Rieske, Ubiquinol--citocromo c reductasa hierro-azufre subunidad	0.818128
O33309	Posible proteína de membrana	0.818991
P9WH95	Proteína ribosómica 50S L33 2	0.819216
P9WJV7	Probable proteína transportadora MmpL2	0.819452
P9WL75	Sec translocon subunidad del complejo accesorio YajC	0.819841
O06546	Probable proteína de transporte transmembrana conservada MmpL13b	0.821764
P9WJV1	Exportador de sideróforos MmpL5	0.822570
I6YFL7	rubredoxina	0.823332
O06568	Proteína no caracterizada	0.823526
P9WM69	Proteína no caracterizada Rv0093c	0.824001

L0T5T4	proteína de la familia PE PE10	0.824556
P9WKV7	Proteína no caracterizada Rv0479c	0.825434
L7N692	Proteína conservada	0.825904
O69663	Posible proteína de membrana conservada	0.826403
O53417	Proteína conservada	0.826960
P9WIR1	Antígeno proteico de bajo peso molecular 6 , CFP-6	0.828839
P9WPU1	ATPasa tipo P exportadora de cobre, ATPasa A tipo P exportadora de cobre	0.828925
O07422	Proteína de membrana asociada a Mce probablemente conservada	0.829488
P71707	Proteína de unión a penicilina 1A , PBP-1A, proteína de unión a penicilina 1, transglicosilasa insensible a la penicilina, peptidoglicano TGasa, transpeptidasa sensible a la penicilina, DD-transpeptidasa	0.830087
P96400	Proteína transmembrana probablemente conservada	0.831052
O07436	L,D-transpeptidasa 4 , LDT 4, Ldt(Mt4)	0.831884
P9WJC3	Proteína asociada a la secreción de ESX-1 EspJ, TB27.4	0.833429
P9WJV3	Exportador de sideróforos MmpL4	0.833660
Q10690	Proteína no caracterizada Rv2082	0.833696
L7N673	Proteína no caracterizada	0.834613
O06213	Proteína de membrana rica en prolina probablemente conservada	0.835225
O50464	Posible proteína de membrana	0.835599
P9WIE7	Proteína de la familia PE-PGRS no caracterizada PE_PGRS46	0.835749
P9WJU3	Probable proteína transportadora MmpL9	0.836156
P9WN55	Proteína H del sistema de escisión de glicina	0.836806
Q6MX50	Proteína de la familia PE-PGRS PE_PGRS5	0.839125
O53536	Proteína que contiene el dominio CN hidrolasa	0.841468
I6YG92	Proteína de membrana probablemente conservada	0.842912
P9WHN5	Pup de proteína procariótica similar a la ubiquitina , modificador bacteriano similar a la ubiquitina	0.843604
P9WI29	Proteína de la familia PPE no caracterizada PPE16	0.843783
P9WM27	Proteína no caracterizada Rv1322	0.843940
I6Y8X8	transposasa putativa	0.844930
Q79FS5	Proteína de la familia PE-PGRS PE_PGRS22	0.845105
I6X7Z3	Proteína transmembrana probablemente conservada	0.846205
L7N6A1	Proteína de la familia PE-PGRS PE_PGRS38	0.846758

P9WQ33	Proteína no caracterizada Rv1318c	0.849465
P9WGF5	Probable proteína del sistema de salida de cationes Rv2025c	0.849593
O06272	Proteína hipotética rica en arginina y prolina	0.849738
L0T2W6	Proteína de la familia Mce Mce1F	0.850122
L0T9U5	Proteína del fago PhiRv1 probable	0.850527
P96878	Anhidrasa carbónica	0.850549
O06312	Transportador de magnesio MgtE	0.851439
Q79FU0	Proteína de la familia PE-PGRS PE_PGRS18	0.852048
P9WPT1	ATPasa transportadora de calcio CtpE	0.853575
P9WPG7	Fosfatidilinositol fosfato sintasa , PIP sintasa, CDP-diacilglicerol--D-mio-inositol-3-fosfato 3-fosfatidiltransferasa	0.854360
P9WJV9	Probable proteína de transporte MmpL1	0.854694
O05574	Proteína rica en serina conservada	0.854831
P9WNK5	Proteína similar a ESAT-6 EsxB, antígeno filtrado de cultivo de 10 kDa CFP-10, CFP-10, proteína antigénica secretada MTSA-10	0.855303
P9WJU7	Exportador de dimicocerosato de ftiocerol MmpL7	0.856000
O65934	Transportador ABC proteína de unión a ATP/permeasa Rv1747	0.857001
I6YA75	Proteína transmembrana probablemente conservada	0.858173
P9WLL1	Proteína no caracterizada Rv2077c	0.858215
P9WNL7	Probable arabinosyltransferase B	0.859102
P9WNL5	Probable arabinosyltransferase C	0.859912
P9WQ29	Proteína no caracterizada Rv1320c	0.862090
Q79FC3	Proteína de la familia PE-PGRS PE_PGRS45	0.862415
P95021	Proteína no caracterizada	0.863365
P71906	Probable proteína transmembrana de transporte de dicarboxilato C4 DctA	0.864463
Q6MX04	Proteína de la familia PPE PPE53	0.867911
O53167	Posible proteína de membrana	0.868665
O53896	Serina proteasa PepD , serina proteasa similar a HtrA	0.868925
O07424	Proteína transmembrana probablemente conservada	0.870196
P9WMT9	Factor de elongación de la transcripción GreA, factor de escisión de transcripción GreA	0.870238
Q6MWW8	Proteína de la familia PE-PGRS PE_PGRS55	0.873442
P9WJ71	Anti-sigma-D factor RsdA , Regulador de SigD, Sigma-D anti-sigma factor RsdA	0.874275

P95115	Posible proteína secretada conservada	0.875306
P9WLI5	Proteína no caracterizada Rv2206	0.875444
L0T8G6	Proteína no caracterizada	0.877072
O53857	Sensor histidina quinasa NarS	0.877131
I6WZ58	Proteína transmembrana probablemente conservada	0.877735
P9WG03	Proteína permeasa del sistema de transporte de trehalosa SugA	0.878871
I6XFB7	probable proteína de membrana	0.879466
Q79FD4	Proteína de la familia PE-PGRS PE_PGRS43	0.880736
P9WPI7	Proteína no caracterizada Rv2325c	0.882606
I6YCQ7	Posible proteína de membrana	0.882607
O53553	Proteína de la familia PE-PGRS no caracterizada PE_PGRS54	0.883110
P9WPS9	Probable ATPasa F transportadora de cationes	0.883111
O69637	Proteína transmembrana probablemente conservada	0.883688
P9WI71	Serina/treonina-proteína quinasa PknH	0.883898
P94973	Proteína de membrana hipotética conservada	0.883953
P9WQ31	Proteína no caracterizada Rv1319c	0.884160
I6Y460	Proteína transmembrana probablemente conservada	0.884457
P9WJN5	Canal mecanosensible de gran conductancia	0.884848
P9WG19	Proteína de división celular FtsX	0.884979
O06583	Proteína de membrana conservada	0.886305
P9WP11	Quinona reductasa Rv1558 dependiente de F420H(2), Fqr	0.888360
O06230	Posible proteína transmembrana conservada	0.888514
O53487	Proteína no caracterizada	0.889344
P9WJV5	Exportador de monomicolato de trehalosa MmpL3, exportador TMM MmpL3, transportador MmpL3, proteína de membrana micobacteriana grande 3	0.889582
P9WMA3	Proteína no caracterizada Rv0010c	0.889911
P9WI53	Decaprenilfosforil-5-fosforribosa fosfatasa putativa Rv3807c , DPPR fosfatasa, Fosfolípido fosfatasa	0.892155
O06625	Proteína de membrana hipotética conservada	0.893295
P9WIF9	Proteína de la familia PE-PGRS no caracterizada PE_PGRS20	0.893300
O53316	Proteína que contiene el dominio DUF4129	0.893316
P9WHS5	Homólogo de proteasa HtpX	0.893318
O06278	Proteína transmembrana probablemente conservada rica en alanina y arginina y prolina	0.895372

P9WK93	Fosfatidilglicerol--prolipoproteína diacilgliceril transferasa	0.896427
O05884	Proteína de transporte transmembrana probablemente conservada	0.896934
O53315	Proteína no caracterizada	0.897997
Q79FP2	Proteína de la familia PE-PGRS PE_PGRS27	0.899270
P9WGP1	Subunidad de proteína translocasa SecD	0.899743
O06827	Proteína de membrana conservada	0.900375
O53582	Beta-(1->2)-arabinofuranosiltransferasa terminal	0.901092
O53633	Posible aciltransferasa transmembrana	0.901123
O05579	Proteína transmembrana probablemente conservada	0.903186
P9WQ35	Adenilato ciclasa , ATP pirofosfato-liasa, Adenilil ciclasa	0.903282
O06252	Proteína transmembrana probablemente conservada	0.903302
P9WJK3	Probable biosíntesis de peptidoglicano proteína MviN	0.903480
P9WKU3	Proteína no caracterizada Rv0497	0.904331
P9WPG3	Probable fosfatidilglicerofosfato sintasa , PGP sintasa, PGPS, CDP-diacilglicerol--glicerol-3-fosfato fosfatidiltransferasa	0.904551
I6Y7L4	Proteína de la familia PPE PPE8	0.905069
P96882	Proteína transmembrana probablemente conservada	0.905085
I6X552	Probable proteína transmembrana rica en alanina conservada	0.905095
P9WL89	Proteína no caracterizada Rv2571c	0.906320
P96883	Proteína transmembrana probablemente conservada	0.906791
P96235	Proteína transmembrana probablemente conservada	0.907254
O53781	Proteína de membrana integral probablemente conservada	0.907750
P9WGN9	Subunidad de proteína translocasa SecF	0.908785
P9WGN5	Probable proteína de membrana exportadora de proteínas SecG	0.910599
P9WL13	Proteína no caracterizada Rv2929	0.912363
P96220	Posible proteína de membrana	0.912591
I6YGV9	Probable proteína de membrana integral transportadora de dipéptidos Transportador ABC DppB	0.913331
P9WPG1	CDP-diacilglicerol--serina O-fosfatidiltransferasa , fosfatidilserina sintasa	0.914099
P9WMA7	Proteína no caracterizada Rv0007	0.914211
P9WN05	Probable dolichyl-fosfato-manosa-proteína manosiltransferasa	0.915579
O53448	Proteína conservada	0.916012
O33358	Posible proteína transmembrana conservada	0.916568
P71708	Proteína transmembrana probablemente conservada	0.916582

P9WFF9	Undecaprenil-difosfatasa , Proteína de resistencia a la bacitracina, Undecaprenil pirofosfato fosfatasa	0.916600
I6Y3P5	Proteína de membrana integral conservada YrbE4B. Posible transportador ABC	0.916910
O69723	Posible transporte de osmoprotectores (glicina betaína/carnitina/colina/L-prolina) Proteína integral de membrana transportadora ABC ProW	0.916941
O53431	Proteína no caracterizada	0.918001
O53713	Proteína rica en glicina hipotética conservada	0.918108
P9WG05	Sistema de transporte de fosfato proteína permeasa PstC 2	0.919246
P9WKW1	Proteína no caracterizada Rv0476	0.919575
O69704	Exportador probable de treonina/serina	0.919585
O53965	Proteína de membrana integral hipotética conservada YrbE3A	0.919800
P96418	Proteína transmembrana probablemente conservada	0.920268
L0T2Q9	Proteína de membrana integral conservada YrbE1B	0.920400
P9WLI7	Proteína no caracterizada Rv2203	0.922751
P9WL85	Proteína no caracterizada Rv2575	0.923167
P9WP07	Proteína de membrana no caracterizada Rv2637	0.923223
P9WHR1	Putativa metaloproteasa de zinc Rip3	0.923724
P96229	Proteína de la familia GDT1	0.923738
P9WIT5	Insertasa de proteína de membrana YidC , Foldasa YidC, Integrasa de membrana YidC, Proteína de membrana YidC	0.924322
I6Y4C4	Proteína que contiene el dominio DUF4185	0.924423
P96266	Posible proteína transmembrana	0.925504
I6Y870	Proteína de membrana integral hipotética conservada YrbE2A	0.926149
O06311	Proteína de membrana probablemente conservada	0.926304
I6Y1H7	Probable proteína integral conservada de membrana rica en alanina y valina y leucina	0.926360
O06397	Posible proteína de membrana conservada	0.926611
O53503	Proteína transmembrana probablemente conservada	0.926798
O07744	Posible proteína transmembrana	0.927296
Q6MX28	Proteína de la familia PE-PGRS PE_PGRS7	0.927393
O53966	Proteína de membrana integral hipotética conservada YrbE3B	0.927475
O53495	Probable proteína transmembrana FxsA	0.927611
P9WG11	Sistema de transporte de fosfato proteína permeasa PstA 1	0.927752
P9WGF7	Probable transportador de sulfato Rv1739c	0.927954

I6XER9	probable proteína transmembrana	0.928207
I6YHB0	Posible proteína similar a histonas Hns	0.928219
O33223	probable proteína transmembrana	0.928326
I6XWF9	Proteína no caracterizada	0.929337
O50377	Proteína transmembrana conservada	0.929857
I6X824	Proteína no caracterizada	0.930127
O53222	Proteína no caracterizada	0.930722
P9WPV7	ATP sintasa subunidad a , ATP sintasa sector F0 subunidad a, F-ATPasa subunidad 6	0.931151
L7N685	Proteína no caracterizada	0.931310
P9WFM5	Putativa proteína de transporte Rv0205	0.932264
O53385	Proteína de membrana integral probablemente conservada	0.932414
P9WFM3	Putativa proteína de transporte Rv1101c	0.932630
I6X7J6	Proteína de membrana probablemente conservada	0.933006
P71896	Probable proteína de membrana integral transportadora de azúcar transportador ABC UspA	0.933352
I6YG51	Posible proteína transmembrana conservada	0.933689
Q79FC4	Proteína no caracterizada	0.934307
O53514	Proteína de membrana integral probablemente conservada	0.934392
O50451	probable proteína transmembrana	0.934900
O53572	Proteína de membrana probablemente conservada	0.936417
I6YBN6	Posible transporte de Mg ²⁺ tipo P ATPasa C MgtC	0.936482
I6YA25	Proteína transmembrana probablemente conservada	0.937221
P95172	NADH-quinona oxidoreductasa subunidad J	0.937422
P9WPD7	Transportador no caracterizado Rv2685	0.938195
P9WLI9	Proteína no caracterizada Rv2197c	0.938397
P96876	Proteína de membrana integral probablemente conservada	0.938444
O06128	Posible proteína de membrana conservada	0.938911
P71729	Proteína que contiene dominio de competencia	0.939331
O86314	probable proteína de membrana	0.939406
O07241	Posible proteína de membrana conservada	0.939465
O06410	Posible proteína transmembrana conservada	0.939492
P9WM01	Proteína no caracterizada Rv1362c	0.939834
P9WMZ7	Alfa-(1->3)-arabinofuranosiltransferasa, Arabinofuranan 3-O-arabinosiltransferasa, Arabinofuranosiltransferasa C	0.939956

O69662	Posible proteína transmembrana conservada	0.939984
O33248	Proteína de membrana integral probable	0.940216
O06790	Proteína conservada	0.940330
P9WLX1	Proteína no caracterizada Rv1490	0.940540
L7N651	Proteína no caracterizada	0.940964
P9WG97	Proteína translocasa proteína independiente de sec TatC	0.941661
I6X849	Proteína rica en prolina hipotética conservada	0.941953
P9WGA7	Proteína similar a SURF1 no caracterizada Rv2235	0.942155
O06404	Posible proteína de membrana conservada	0.942438
P95017	Proteína transmembrana probablemente conservada	0.942574
P9WPG5	Putativa cardiolipina sintasa	0.944016
P9WJW7	Transportador de tipo MFS no caracterizado Rv2994	0.944021
P9WHW1	Probable peptidil-prolil cis-trans isomerasa B , PPIasa B, Rotamasa B	0.944028
P96886	Subunidad epsilon de carboxilasa A de acetil-/propionil-coenzima dependiente de biotina	0.944343
P71787	probable proteína de membrana	0.944343
P9WFR5	Decaprenil-fosfato fosforribosiltransferasa , 5-fosfo-alfa-D-ribosa-1-difosfato:decaprenil-fosfato 5-fosforribosiltransferasa, DPPR sintasa	0.944450
O53150	Alfa-(1->6)-manopiranosiltransferasa Rv1459c ,	0.944874
P9WFS5	Proteína de membrana de la familia TVP38/TMEM64 Rv0625c	0.945892
P9WNP9	Proteína EccD5 del sistema de secreción ESX-5 , componente D5 conservado de ESX , proteína EccD5 del sistema de secreción tipo VII, proteína EccD5 de T7SS	0.946030
O69722	Posible transporte de osmoprotectores (Glicina betaína/carnitina/colina/L-prolina) proteína de membrana integral transportador ABC ProZ	0.946475
P9WLS5	Proteína de membrana no caracterizada Rv1735c	0.946728
P9WN03	Galactano 5-O-arabinofuranosiltransferasa , Arabinofuranosiltransferasa AftA	0.946757
P71745	Sistema de transporte de sulfato proteína permeasa CysT	0.946760
P71796	Proteína transmembrana hipotética conservada	0.947040
P96834	NAD(P) transhidrogenasa subunidad beta , nicotinamida nucleótido transhidrogenasa subunidad beta	0.947891
P9WG09	Sistema de transporte de fosfato proteína permeasa PstA 2	0.948096
P9WPD9	Transportador no caracterizado Rv2684	0.948112

O06562	Probable nitrato reductasa respiratoria (cadena gamma) NarI	0.948407
O33237	Proteína transmembrana probablemente conservada	0.948641
O86351	Proteína no caracterizada	0.948841
P9WP37	Citocromo bc1 complejo citocromo b subunidad , Citocromo bc1 reductasa complejo subunidad QcrB, Ubiquinol--citocromo c reductasa citocromo b subunidad	0.948962
O06320	Posible proteína transmembrana conservada	0.949184
O69726	Proteína de membrana no caracterizada Rv3760	0.949231
O53368	Subunidad de membrana de succinato deshidrogenasa 2 SdhC	0.949960
O53731	Posible proteína transmembrana conservada	0.950029
O53483	Probable transportador ABC de proteína de membrana integral de transporte de azúcar	0.950488
I6YE67	Probable proteína transmembrana rica en alanina y leucina conservada	0.951003
P9WKR5	Proteína no caracterizada Rv0876c	0.951380
I6YAV3	Proteína no caracterizada	0.951853
O69709	Posible oxidorreductasa	0.951945
P71590	Proteína que contiene el dominio FHA FhaA	0.952060
P72045	Probable proteína transmembrana conservada rica en alanina y leucina	0.952070
O06539	Proteína de membrana hipotética conservada	0.952282
P9WG13	Sistema de transporte de molibdeno proteína permeasa ModB	0.952633
P72049	Probable transporte de antígeno O/lipopolisacárido proteína de membrana integral transportador ABC RfbD	0.952880
I6X961	probable proteína transmembrana	0.952892
I6YC44	Posible proteína transmembrana conservada	0.953175
P9WNQ5	Proteína del sistema de secreción ESX-2 eccD2 , componente conservado ESX D2, proteína del sistema de secreción Tipo VII eccD2, proteína T7SS eccD2	0.953483
O06152	Probable manosiltransferasa. Proteína transmembrana probablemente conservada	0.953588
P9WG21	Resistencia probable a la doxorrubicina Proteína transportadora ABC permeasa DrrC	0.953617
O53546	Proteína de membrana integral conservada YrbE4A. Posible transportador ABC	0.953627
P71746	Probable proteína de membrana integral transportadora de sulfato ABC transportador CysW	0.953649
O06407	Proteína de membrana integral probablemente conservada	0.953684

L7N662	Probable citocromo D de membrana integral ubiquinol oxidasa (Subunidad I) CydA (Citocromo BD-I oxidasa subunidad I)	0.953795
O05317	Proteína de membrana integral probablemente conservada	0.953907
I6YGT7	Posible proteína de membrana conservada	0.953927
P9WG01	Proteína permeasa del sistema de transporte de trehalosa SugB	0.954015
P9WJI3	Intercambiador de Na(+)/H(+) no caracterizado Rv2287	0.954143
P9WP47	Transportador de péptidos CstA , homólogo de proteína A de inanición de carbono	0.954174
O33206	Proteína transmembrana probablemente conservada	0.954210
P9WK99	Lipoproteína señal peptidasa , Prolipoproteína señal peptidasa, Señal peptidasa II, SPasa II	0.954291
O69690	Coordinador de captación de lípidos A	0.954296
I6XFF3	Probable transporte de Sn-glicerol-3-fosfato proteína de membrana integral transportador ABC UgpA	0.954361
P9WLL3	Proteína no caracterizada Rv2076c	0.954393
O07412	Proteína de membrana integral conservada YrbE1A	0.954581
P9WG85	Transportador de tipo MFS no caracterizado Rv1877	0.954711
P9WIA5	Putativa permeasa de fosfato Rv2281	0.955032
O53918	probable proteína de membrana	0.955078
O07790	Proteína de membrana integral hipotética conservada YrbE2B	0.955120
O05895	Probable alcano 1-monooxigenasa transmembrana AlkB (Alcano 1-hidroxilasa) (Ácido láurico omega-hidroxilasa) (Omega-hidroxilasa) (Ácido graso omega-hidroxilasa) (Alcano hidroxilasa-rubredoxina)	0.955204
P9WKR1	Proteína no caracterizada Rv0879c	0.955348
P9WPF7	Fosfatidato citidililtransferasa , CDP-DAG sintasa, CDP-DG sintasa, CDP-diacilglicerol sintasa, CDS, CDP-diglicérido pirofosforilasa, CDP-diglicérido sintasa, CTP:fosfatidato citidililtransferasa	0.955363
P9WNB7	Fumarato reductasa subunidad C , Quinol-fumarato reductasa subunidad C, QFR subunidad C	0.955436
I6X7F2	Posible proteína exportada	0.955493
I6YET7	Permeasa putativa Rv2963	0.955619
P9WJG7	Proteína de membrana no caracterizada ArfB	0.955746
I6X4W4	Probable proteína de membrana integral de transporte de arsénico ArsC	0.955773
L0T550	Putativa metaloproteasa de zinc Rip2	0.955787
P9WPR7	Transportador no caracterizado Rv0917	0.955925

P9WLB1	Putativa proteína secretada Rv0236.1	0.955997
P9WLA1	Proteína no caracterizada Rv2560	0.956256
P9WFS3	Proteína de membrana de la familia TVP38/TMEM64 Rv1491c	0.956382
P9WNQ3	Proteína EccD3 del sistema de secreción ESX-3 , componente D3 conservado de ESX , proteína EccD3 del sistema de secreción tipo VII, proteína EccD3 de T7SS	0.956929
P9WG07	Sistema de transporte de fosfato proteína permeasa PstC 1	0.956938
O05883	Metanotiol S-metiltransferasa	0.956968
I6Y1U3	Probable transporte de Sn-glicerol-3-fosfato proteína de membrana integral transportador ABC UgpE	0.957313
I6Y8Z3	Posible proteína de membrana integral conservada	0.957318
I6YEL8	Proteína de membrana integral probablemente conservada	0.957453
P9WM31	Proteína no caracterizada Rv1303	0.957747
O33359	Posible proteína transmembrana	0.957906
P9WMZ9	Alfa-(1-2)-fosfatidilinositol manósido manosiltransferasa dependiente de poliprenol-fosfato-manosa , alfa-D-manosa-alfa-(1-2)-manosiltransferasa, alfa-manosiltransferasa, alfa-ManT, PPM -manosiltransferasa dependiente, poliprenol-fosfato-manosa alfa-manosiltransferasa, PPM alfa-manosiltransferasa	0.957950
P9WG23	Resistencia a la doxorrubicina Proteína transportadora ABC permeasa DrrB	0.958015
P9WP45	Probable polipéptido 4 de citocromo c oxidasa , subunidad 4 del citocromo aa3, polipéptido IV de citocromo c oxidasa	0.958218
I6Y748	Proteína transmembrana probablemente conservada	0.958366
P9WIW3	Proteína no caracterizada Rv0083	0.958862
O05318	Proteína permeasa del sistema de salida de múltiples fármacos Rv1217c	0.959017
P9WLW1	Proteína no caracterizada Rv1510	0.959191
I6YGY9	Proteína rica en prolina hipotética conservada	0.959211
P9WIZ5	Transportador de cationes metálicos divalentes MntH , Mramp	0.959298
O86349	Probable transportador ABC de membrana integral de transporte de antibióticos no identificado	0.959396
P9WG87	Transportador de tipo MFS no caracterizado Rv1250	0.959675
P95289	Proteína no caracterizada	0.959687
I6YB21	Proteína no caracterizada Rv3067	0.960378
O65935	Posible proteína de membrana integral	0.960408
P9WN97	Probable peptidoglicano glicosiltransferasa FtsW , PGT, proteína de división celular FtsW, polimerasa de pared celular, peptidoglicano polimerasa, PG polimerasa	0.960412

I6Y9M6	Proteína que contiene el dominio Peptidase_A24	0.960630
O06132	probable proteína de membrana	0.960648
P9WLT5	Proteína no caracterizada Rv1591	0.960707
P9WJX1	Transportador de tipo MFS no caracterizado Rv2456c	0.960709
P9WJB3	Las fluoroquinolonas exportan proteína permeasa Rv2686c	0.960933
P9WLG3	Proteína no caracterizada Rv2265	0.960939
P96247	Proteína no caracterizada	0.960984
P9WJY1	Transportador de tipo MFS no caracterizado Rv0037c	0.960988
P9WP57	Proteína de división celular CrgA	0.961330
I6YEJ7	Sistema de salida de níquel/cobalto	0.961548
P9WIP3	1,4-dihidroxi-2-naftoato octapreniltransferasa , DHNA-octapreniltransferasa, isoprenil difosfato: 1,4-dihidroxi-2-naftoato isopreniltransferasa	0.961656
P96410	Aciltransferasa de membrana integral probable	0.961723
P9WNQ7	Proteína EccD1 del sistema de secreción ESX-1 , componente D1 conservado de ESX, proteína secretora Snm4, proteína EccD1 del sistema de secreción tipo VII, proteína EccD1 de T7SS	0.961999
O33203	Probable proteína transportadora de D-serina/alanina/glicina CycA	0.962043
I6XWB9	Proteína no caracterizada	0.962312
P9WKN3	Proteína no caracterizada Rv0955	0.962583
P71916	Proteína no caracterizada	0.962856
P96820	Proteína transmembrana probablemente conservada	0.962889
O33362	Posible aciltransferasa de membrana	0.963127
Q10881	Posible formiato hidrogenoliasa HycD (FHL)	0.963212
O06540	Proteína de membrana integral probable	0.963347
P9WK33	Transportador putativo de aminoácidos Rv0488	0.963442
I6YCG9	Posible bomba de arsénico proteína de membrana integral ArsB2	0.963617
P9WM63	Proteína no caracterizada Rv0102	0.963748
O33188	Proteína permeasa de transporte	0.963783
O06139	Probable citocromo D de membrana integral ubiquinol oxidasa (Subunidad II) CydB (Citocromo BD-I oxidasa subunidad II)	0.963818
O05896	Posible proteína de membrana integral de transporte catiónico de aminoácidos	0.963830
P9WKM9	Proteína no caracterizada Rv0961	0.963968
L7N665	Proteína de transporte de membrana integral conservada probable	0.964236

P96833	NAD(P)(+) transhidrogenasa translocadora de protones	0.964338
P9WNQ1	Proteína del sistema de secreción ESX-4 eccD4 , componente conservado ESX D4, proteína del sistema de secreción Tipo VII eccD4, proteína T7SS eccD4	0.964473
P9WQ65	Transportador de amonio , canal de amoníaco, canal de amonio	0.964742
O53445	Probable proteína de membrana rica en glicina y prolina	0.964790
O53508	Alfa-(1->6)-manopiranosiltransferasa A , MptA, alfa-(1->6)-manopiranosiltransferasa dependiente de poliprenil-1-monofosforilmanosa, alfa-(1->6)-manopiranosiltransferasa dependiente de PPM	0.964929
P9WK31	Exportador de lisina LysE	0.964961
P9WKX9	Proteína no caracterizada Rv3630	0.964988
O07235	Proteína de membrana integral probablemente conservada	0.965208
P9WN01	Alfa-(1-2)-fosfatidilinositol pentamanósido manosiltransferasa dependiente de poliprenol-fosfato-manosa , alfa-manosiltransferasa, Alpha-ManT, manosiltransferasa dependiente de PPM, alfa-manosiltransferasa de poliprenol-fosfato-manosa, alfa-manosiltransferasa de poliprenol-fosfato-manosa, PPM alfa- manosiltransferasa	0.965309
O07730	Probable proteína de membrana integral de transporte de ácido siálico NanT	0.965506
O33262	Proteína transmembrana conservada	0.965585
I6Y9U6	Proteína transmembrana probablemente conservada	0.965616
O53484	Probable transportador ABC de proteína de membrana integral de transporte de azúcar	0.965718
P9WG93	Proteína de membrana no caracterizada Rv2723	0.965813
O53369	Probable succinato deshidrogenasa (Subunidad de anclaje de membrana hidrofóbica) SdhD (Succínico deshidrogenasa) (Fumarato reductasa) (Fumarato deshidrogenasa) (Fumarica hidrogenasa)	0.965991
Q11064	Probable aciltransferasa	0.966313
O07802	Savia exportadora de sulfolípidos-1, proteína de direccionamiento de sulfolípidos-1	0.966554
O06170	Posible proteína de membrana rica en prolina conservada	0.966676
P9WP09	Proteína de membrana no caracterizada Rv0364	0.966705
O06393	Posible proteína de biogénesis tipo citocromo C CcsA	0.966870
I6XHC3	Proteína de membrana integral probable	0.966941
O33234	Probable proteína integral conservada de membrana rica en alanina, valina y leucina	0.966975
P9WM19	Proteína no caracterizada Rv1342c	0.966983

O53925	Proteína de membrana integral probable	0.967127
P9WP71	Probable subunidad 1 de citocromo c oxidasa , subunidad 1 de citocromo aa3, polipéptido I de citocromo c oxidasa	0.967222
L0TDU1	Probable proteína de membrana integral de transporte de azúcar SugI	0.967228
P9WIX1	NADH-quinona oxidoreductasa subunidad H , NADH deshidrogenasa I subunidad H, NDH-1 subunidad H	0.967236
P95152	Proteína de membrana integral probablemente conservada	0.967512
P9WIM7	Homólogo de antígeno rico en prolina , Pra	0.967715
P71892	Proteína de membrana integral de transporte de aminoácidos catiónicos probables RocE	0.967775
P9WLI3	Proteína no caracterizada Rv2209	0.967813
P96894	Posible proteína transmembrana	0.968181
P9WLF5	Proteína no caracterizada Rv2272	0.968241
P9WQM5	Transportador no caracterizado Rv1979c	0.968306
P9WIW5	NADH-quinona oxidoreductasa subunidad M , NADH deshidrogenasa I subunidad M, NDH-1 subunidad M	0.968531
P9WG89	Homólogo de proteína B de resistencia a múltiples fármacos	0.968531
P71616	Posible proteína F DinF inducible por daño en el ADN	0.968559
P9WJT1	Probable proteína accesoria de transporte MmpS3	0.968619
P9WIW1	NADH-quinona oxidoreductasa subunidad L , NADH deshidrogenasa I subunidad L, NDH-1 subunidad L	0.969057
O06141	Proteína de membrana probablemente conservada	0.969278
O53854	Proteína de membrana integral probablemente conservada	0.969408
O07432	Proteína transmembrana probablemente conservada	0.969475
I6XHB8	Probable proteína de membrana integral de transporte de ácido dicarboxílico KgtP (transportador de dicarboxilato)	0.969544
P96805	Proteína transmembrana probablemente conservada	0.969725
P71883	Probable proteína de extrusión de nitrito 1 NarK1 (facilitador de nitrito 1)	0.969744
L7N6B9	Probable GABA permeasa GabP (portador de transporte de 4-aminobutirato) (GAMA-aminobutirato permeasa)	0.969865
P9WIW9	NADH-quinona oxidoreductasa subunidad N , NADH deshidrogenasa I subunidad N, NDH-1 subunidad N	0.969993
Q6MX47	Proteína de membrana integral probablemente conservada	0.970044
O53420	Proteína transmembrana probablemente conservada	0.970280
P9WQM9	L-asparagina permeasa 1 , proteína de transporte de L-asparagina 1	0.970396

P9WJY3	Probable transportador de triacilglicéridos Rv1410c , transportador de salida de fármacos tipo MFS P55	0.970418
P9WG91	Proteína de resistencia a múltiples fármacos Stp , bomba de eflujo de tetraciclina de espectinomina	0.970491
O50384	Proteína de membrana integral probable	0.970595
O05849	Proteína de membrana integral probablemente conservada	0.970618
O06378	Proteína de membrana integral probablemente conservada	0.970801
I6X548	Probable proteína integral conservada de membrana rica en alanina y leucina	0.970927
O07729	Proteína de membrana probablemente conservada	0.971005
P9WM21	Proteína no caracterizada Rv1337	0.971121
O53632	Proteína de membrana integral probablemente conservada	0.971382
P9WIX3	NADH-quinona oxidoreductasa subunidad K , NADH deshidrogenasa I subunidad K, NDH-1 subunidad K	0.971420
O06321	Proteína de membrana integral probablemente conservada	0.971426
P9WJW9	Bomba de salida de drogas JefA	0.971623
P71749	Proteína de membrana integral probablemente conservada	0.971686
P9WMW5	Decaprenil-fosfato N-acetilglucosaminofosfotransferasa , Decaprenil-fosfato GlcNAc-1-fosfato transferasa, Decaprenil-fosfato alfa-N-acetilglucosaminil 1-fosfato transferasa, UDP-GlcNAc:decaprenil-fosfato GlcNAc-1-fosfato transferasa	0.971711
P9WLR7	Proteína no caracterizada Rv1824	0.971751
P9WJX5	Transportador de tipo MFS no caracterizado Rv0849	0.971867
O06781	Proteína de membrana integral probablemente conservada	0.971906
P9WQM7	L-asparagina permeasa 2 , proteína de transporte de L-asparagina 2	0.972134
O53668	Proteína de membrana integral probablemente conservada	0.972155
P95218	Proteína de extrusión de nitrito de membrana integral probable NarK3 (facilitador de nitrito)	0.972187
P71757	Proteína conservada	0.972202
P9WLJ5	Proteína no caracterizada Rv2091c	0.972263
I6Y2H3	Probable hidroxilaminobenceno mutasa Hab	0.972267
O53671	Probable succinato deshidrogenasa [subunidad de anclaje de membrana] (succínico deshidrogenasa)	0.972292
O53919	Proteína de transporte de membrana integral conservada probable	0.972408
O53910	Probable proteína de membrana integral del transportador iónico ChaA	0.972427

P9WN99	Peptidoglicano glicosiltransferasa RodA , transglicosilasa RodA no canónica	0.972441
O06171	Probable proteína rica en alanina y leucina de membrana integral conservada	0.972541
O06403	Proteína de membrana integral probable	0.972700
P9WP67	Probable subunidad 3 de citocromo c oxidasa , subunidad 3 de citocromo aa3, polipéptido III de citocromo c oxidasa	0.972835
P95224	Transportador de nitrato/nitrito	0.972918
P95154	Proteína transmembrana probablemente conservada	0.973029
P9WJY5	Transportador de tipo MFS no caracterizado EfpA , proteína de eflujo A	0.973039
P9WMW7	Fosfo-N-acetilmuramoil-pentapéptido-transferasa , UDP-MurNAc-pentapéptido fosfotransferasa	0.973879
O50427	Proteína no caracterizada	0.974111
P9WJX7	Bomba de eflujo de cloranfenicol Rv0191	0.974237
P9WM75	Proteína no caracterizada Rv0085	0.974447
P9WLF3	Proteína no caracterizada Rv2273	0.974938
P9WP61	Transportador de iones de fluoruro putativo CrcB 2	0.974971
O33255	Proteína de membrana integral probablemente conservada	0.975188
P9WJX3	Probable transportador de eflujo de múltiples fármacos Rv1634	0.975281
P9WMS9	Biosíntesis de arabinogalactano reclutando proteína Rv3789	0.976169
P9WG51	Lisoplasmalogenasa	0.976339
P9WFN7	UPF0073 proteína de membrana Rv1085c	0.976737
P9WGF1	Proteína de resistencia a múltiples fármacos Mmr	0.977513
P9WIR9	homólogo de proteína antigénica de 34 kDa	0.979206

*Valor de probabilidad de ser de secreción (la proteína es de secreción con SecP score > 0.5)

Anexo 5. Proteínas secretadas por *M. tuberculosis* predichas con destino nuclear en la célula huésped (n=58), con sus respectivas propiedades fisicoquímicas (PM y pI) determinados mediante ProtParam.

	Uniprot ID	Nombre de la proteína	PM(KDa)	pI
Proteínas predichas de secreción y con ubicación nuclear (n=48)	P9WIR7	Proteína secretada rica en alanina y prolina Apa	32.7	4.9
	O05458	Mycosin-2 , 3.4.21.-, MycP2 proteasa	55.5	5.6
	P9WIY3	8-oxo-(d)GTP fosfatasa , 8-oxo-(d)GTPasa,	34.7	9.2

	MtuMutT1, proteína mutante MutT1		
O07415	Proteína de la familia Mce Mce1C	53.8	5.3
O06555	Proteína rica en ala conservada	36.4	5.4
Q79FV3	Proteína de la familia PE-PGRS PE_PGRS15	50.2	4
P9WK79	Lipoproteína putativa LppB	24.1	6.1
P9WJC1	Proteína asociada a la secreción de ESX-1 EspK	74.4	5.3
P9WJD9	Proteína asociada a la secreción de ESX-1 EspB , antígeno MTB48	47.5	4.7
P9WMT9	Factor de elongación de la transcripción GreA , Factor de escisión de la transcripción GreA	17.8	4.9
P9WNK5	Proteína similar a ESAT-6 EsxB , antígeno filtrado de cultivo de 10 kDa CFP-10, CFP-10, proteína antigénica secretada MTSA-10	10.7	4.6
I6YHB0	Posible proteína similar a histonas Hns	13.8	10.5
O53572	Proteína de membrana probablemente conservada	27.1	5.8
P9WJC5	Proteína asociada a la secreción de ESX-1 EspI	70.6	9.7
P9WKW5	Proteína de membrana no caracterizada Rv3835	47.1	5.1
O05574	Proteína rica en serina conservada	11.4	9
P9WH95	Proteína ribosómica 50S L33 2	6.4	9.8
O06404	Posible proteína de membrana conservada	55.1	4.4
P9WM27	Proteína no caracterizada Rv1322	11.3	10.7
I6X562	Proteína no caracterizada	8.9	6.7
P9WIR9	homólogo de proteína antigénica de 34 kDa	30.2	6.2
P9WLI5	Proteína no caracterizada Rv2206	26.2	11.6
I6YC95	Proteína de la familia Mce Mce4F	59.6	5.5
P9WJC3	Proteína asociada a la secreción de ESX-1 EspJ	27.3	4.1
O53222	Proteína no caracterizada	9.5	10.2
P95225	Antitoxina putativa Rv0268c	19.4	11.1
O05447	Proteína no caracterizada	10.8	4

	I6YGY9	Proteína rica en prolina hipotética conservada	13	10
	I6YFL7	rubredoxina	6.7	3.9
	I6XWB9	Proteína no caracterizada	6.9	5.2
	I6Y8X8	transposasa putativa	12.3	6.9
	O53401	Proteína no caracterizada	6.7	3.9
	L0T8G6	Proteína no caracterizada	18.8	11.9
	L0T9U5	Proteína del fago PhiRv1 probable	17.9	11.5
	P9WL11	Proteína no caracterizada Rv3073c	13.7	7.9
	O53937	Probable ferredoxina	7.3	4.1
	P9WI49	Proteína de la familia PPE no caracterizada PPE1	46.8	4.7
	P95205	Proteína de la familia de repeticiones 13E12 conservada	13.5	9.9
	O53402	integrasa probable	11.8	10.5
	O06170	Posible proteína de membrana rica en prolina conservada	28.5	7.7
	O06213	Proteína de membrana rica en prolina probablemente conservada	39.6	11.4
	O06790	Proteína conservada	60.6	9.5
	O05883	Metanotiol S-metiltransferasa	28	10
	P96376	Posible proteína de membrana conservada	24.5	9.8
	O53621	probable proteína de membrana	14.6	10.8
	Q10690	Proteína no caracterizada Rv2082	73.5	6.1
	I6X7F9	Regulador transcripcional Rv3488	11.9	9.6
	P9WLF7	Proteína no caracterizada Rv2271	11.5	4.8
Proteínas del secretoma experimental predichas con ubicación nuclear (n=10)	P9WNC3	Formamidopirimidina-DNA glicosilasa 1, Fapy-DNA glicosilasa 1, DNA-(sitioapurínico o apirimidínico) liasa MutM 1, AP liasa MutM 1	31.9	9.9
	A0A089QKZ7	Proteína no caracterizada Rv1155A	6.7	10.3
	P9WG49	ADN topoisomerasa 1, ADN topoisomerasa I, Omega-proteína, Enzima relajante, Swivelasa, Enzima destorcedora	102.3	9.1
	P9WH37	Proteína ribosómica S3 30S	30	10.1

B5TV88	Proteína similar a ESAT-6	10.7	4.5
P9WKN9	Proteína no caracterizada Rv0942	9.7	10.8
P9WFL1	UPF0232 proteína Rv0004	19.9	11.6
I6X642	Proteína no caracterizada	10.1	4.6
P9WMQ1	ADN helicasa dependiente de ATP UvrD1	85	5.3
P9WGF3	Proteína de partición cromosómica Smc	130.6	5.1

Anexo 6: Puntuaciones en distintas escalas obtenidas del refinamiento de las estructuras modeladas de la Formamidopirimidina-ADN glicosilasa (P9WNC3) obtenidos con DeepRefiner.

ESTRUCTURAS MODELADAS	PUNTUACIONES						
	Calidad global	Energía de Rosetta (Kcal/mol)	MolProbity	*GOAP	*OPUS-PSP	*DFIRE	*RWplus
Modelo 1	0.343	-907.405	0.966	-37,010.850	-4,578.651	-480.262	-58,449.493
Modelo 2	0.341	-900.577	1.026	-37,512.770	-4,646.344	-480.602	-58,505.923
Modelo 3	0.340	-903.704	1.026	-37,310.610	-4,586.418	-478.342	-58,199.153
Modelo 4	0.340	-906.538	0.997	-36,970.510	-4,682.844	-479.811	-58,427.659
Modelo 5	0.339	-913.376	0.899	-37,489.260	-4,669.452	-481.227	-58,577.471

*GOAP: Puntuación de energía basada en el potencial estadístico de todos los átomos, dependiente de la orientación generalizada.

*OPUS-PSP: Potencial estadístico de todos los átomos dependiente de la orientación derivada del empaquetamiento de cadena lateral.

*DFIRE: Puntuación de energía derivada del estado de referencia de gas ideal finito.

*RWplus: Puntuación de energía basada en el potencial atómico dependiente de la distancia.

Anexo 7. Puntuaciones en distintas escalas obtenidas del acoplamiento de las conformaciones de los complejos proteína-ADN de I6X7F9 (Regulador transcripcional Rv3488), P9WNC3 (Formamidopirimidina-ADN glicosilasa) y P9WH37 (Proteína ribosómica S3 30S) obtenidos con HDOCK

Proteínas		Complejos modelados									
		1	2	3	4	5	6	7	8	9	10
I6X7F9 (PDB:5ZHC)	Puntuación de acoplamiento	-278.11	-260.00	-250.78	-243.75	-242.12	-239.82	-238.12	-237.65	-236.49	-235.14

P9WNC3	nto (kcal/mol)	-571.27	-282.56	-260.02	-259.56	-256.60	-246.02	-244.04	-243.08	-240.98	-240.69
P9WH37 (PDB: 7KGB)		-306.39	-280.42	-266.42	-263.85	-261.94	-255.34	-250.33	-249.79	-247.91	-242.94

Anexo 8. Tipos de interacciones establecidas entre los residuos de la proteína Ribosómica 30S S3 (P9WH37) con ARN, identificadas en el complejo P9WH37-ARN, computadas con PLIP. El ligando es ARN.

Interacciones catiónicas-π					
Residuo	AA	Distancia		Grupo ligando	Átomo del ligando
142c	ARG	4.02		Aromático	18231942, 1943, 1945, 1946, 1948, 1949
142c	ARG	5.58		Aromático	1962, 1963, 1964, 1965, 1971
Enlaces de hidrógeno					
Residuo	AA	*Distancia H-A	*Distancia D-A	Átomo del ligando	Átomo de la proteína
71 c	ARG	3.18	3.98	577[Ng+]	2007 [Nar]
71c	ARG	2.91	3.73	578[Ng+]	2006[O2]
107c	ASN	3.52	3.89	858[Nam]	1888[O2]
121c	GLU	2.07	2.78	2002[O3]	956[O2]
125c	ASN	2.26	3.12	988[Nam]	2013[O3]
135	LYS	3.01	3.87	1077[N3]	1978[O3]
139	SER	1.55	2.50	1105[O3]	1967[Npl]
142	ARG	2.94	3.67	1128[Ng+]	1936[O3]
142	ARG	3.06	3.76	1129[Ng+]	1936[O3]
143	GLN	2.88	3.76	1947[Npl]	1137[O2]
143	GLN	1.51	2.36	1138[Nam]	1945[Nar]
Puentes salinos					
Residuo	AA	Distancia		Grupo del ligando	Átomo del ligando
48c	ARG	4.42		Fosfato	2054, 2054, 2057, 2056, 2055, 2040
106c	LYS	5.38		Fosfato	1865, 1865, 1866, 1867, 1868, 1853

126c	ARG	5.26	Fosfato	1694, 1694, 1697, 1695, 1696
------	-----	------	---------	------------------------------

* **Distancia H-A:** Distancia entre el hidrógeno y los átomos aceptores

* **Distancia D-A:** Distancia entre átomos donantes y aceptores

* **[Ng+]:** Nitrógeno de guanina

* **[Nam]:** Nitrógeno de amida desprotonada

* **[O3]:** *sp* 3-oxígeno hibridado

* **[Npl]:** Nitrógeno con hibridación *sp* 2 pero sin doble enlace (planar)

* **[O2]:** Oxígeno hibridado *sp* 2

* **[Nar]:** Nitrógeno aromático

Anexo 9. Tipos de interacciones establecidas entre la reguladora transcripcional Rv3488 con el ADN, identificadas en el complejo I6X7F9-ADN, computadas con PLIP. El ligando es ADN de doble hebra.

Interacciones hidrofóbicas					
Residuo	AA	Distancia		Átomo del ligando	Átomo de la proteína
44A	TYR	3.78		1823	391
45A	PRO	3.77		2519	404
Enlaces de hidrógeno					
Residuo	AA	Distancia H-A	Distancia D-A	Átomo del ligando	Átomo de la proteína
24A	GLY	2.15	2.85	232 [Nam]	1785 [O2]
25A	ALA	2.94	3.74	236 [Nam]	1785 [O2]
66A	ARG	3.15	3.79	570 [Ng+]	1750 [O3]
69A	ARG	2.71	3.29	587 [Nam]	1784 [O3]
71A	TYR	3.15	3.70	616 [O3]	1806 [O3]
6A	ARG	2.40	3.31	2482 [O3]	97 [Nox]
Puentes salinos					
Residuo	AA	Distancia		Grupo del ligando	Átomo del ligando
6B	ARG	5.38		Fosfato	1991, 1991, 1979, 1994, 1992, 1993
49B	ARG	5.16		Fosfato	1971, 1971, 1972, 1973, 1974, 1959

68A	ARG	4.63	Fosfato	1763, 1763, 1764, 1765, 1750, 1766
69A	ARG	4.86	Fosfato	1805, 1805, 1791, 1807, 1806, 1808
2B	ARG	3.54	Fosfato	2460, 2460, 2461, 2462, 2463, 2449
2B	ARG	5.11	Fosfato	2441, 2441, 2428, 2443, 2444, 2442
6A	ARG	2.16	Fosfato	2480, 2480, 2483, 2468
66B	ARG	4.35	Fosfato	2235, 2235, 2236, 2237, 2238, 2223
69B	ARG	5.41	Fosfato	2277, 2277, 2264, 2278, 2279, 2280
69B	ARG	05.08	Fosfato	2256, 2256, 2243, 2259, 2257, 2258

* **Distancia H-A:** Distancia entre el hidrógeno y los átomos aceptores

***Distancia D-A:** Distancia entre átomos donantes y aceptores

*[Nam]: Nitrógeno de amida desprotonada

*[Ng+]: Nitrógeno de guanina

*[O3]: *sp* 3-oxígeno hibridado

*[O2]: Oxígeno hibridado *sp* 2

*[Nox]: N-óxido de nitrógeno

Anexo 10. Tipos de interacciones entre los residuos de interfaz de la Formamidopirimidina-ADN glicosilasa(P9WNC3) con el ADN, identificadas en el complejo P9WNC3-ADN, computadas con PLIP. El ligando es ADN de doble hebra.

Interacciones hidrofóbicas					
Residuo	AA	Distancia		Átomo del ligando	Átomo de la proteína
121A	PHE	3.67		2318	917
121A	PHE	3.88		2311	918
237A	VAL	3.19		2395	1803
274A	ASN	3.63		2272	2101
Enlaces de hidrógeno					
Residuo	AA	Distancia H-A	Distancia D-A	Átomo del ligando	Átomo de la proteína

135A	SER	2.59	03.05	1018 [O3]	2624 [O2]
137A	VAL	3.45	04.03	1028 [Nam]	2624 [O2]
176A	GLN	2.40	2.84	1350 [Nam]	2625 [O3]
236A	TYR	2.72	3.44	2534 [Npl]	1799 [O2]
275A	ARG	3.38	3.97	2114 [Ng+]	2612 [Nar]
119A	ARG	3.48	04.01	901 [Ng+]	2313 [Npl]
119A	ARG	1.99	2.96	902 [Ng+]	2333 [Nar]
236A	TYR	2.63	3.56	1788 [Nam]	2384 [O3]
237A	VAL	2.21	3.18	1800 [Nam]	2400 [O3]
238A	ASN	2.35	3.23	1807 [Nam]	2399 [O3]
239A	VAL	1.98	2.80	1815 [Nam]	2399 [O3]
240A	ASN	3.75	04.09	1822 [Nam]	2400 [O3]
273A	MET	3.17	3.91	2091 [Nam]	2221 [O3]
274A	ASN	2.33	2.89	2104 [Nam]	2244 [O2]
274A	ASN	2.19	2.78	2099 [Nam]	2244 [O2]
234A	SER	2.46	03.07	1777 [O3]	2578 [Nar]
253A	TYR	3.45	04.04	1930 [O3]	2582 [O3]
276A	SER	3.33	04.04	2118 [Nam]	2563 [O2]
165A	LYS	2.30	03.05	1255 [N3]	2204 [O3]
167A	SER	2.51	3.15	1271 [O3]	2204 [O3]
171A	ARG	2.24	2.71	2201 [O3]	1310 [O2]
175A	ASP	3.57	3.93	1341 [O3]	2201 [O3]
3A	GLU	1.76	2.60	2583 [O3]	14 [O3]
119A	ARG	1.94	2.78	901 [Ng+]	2593 [O2]
119A	ARG	2.98	3.60	902 [Ng+]	2593 [O2]
184A	ASN	3.13	04.07	1400 [Nam]	2583 [O3]
275A	ARG	2.63	3.45	2114 [Ng+]	2583 [O3]
275A	ARG	2.17	3.11	2115 [Ng+]	2583 [O3]

Puentes salinos

Residuo	AA	Distancia	Grupo del ligando	Átomo del ligando
35A	ARG	3.95	Fosfato	2644, 2644, 2643, 2645, 2646, 2647
61A	LYS	2.96	Fosfato	2602, 2602, 2603, 2604, 2605

83A	HIS	05.08		Fosfato	2602, 2602, 2603, 2604, 2605
32A	ARG	04.01		Fosfato	2320, 2320, 2321, 2322, 2323, 2319
100A	HIS	4.27		Fosfato	2359, 2359, 2358, 2361, 2362, 2360
171A	ARG	4.31		Fosfato	2220, 2220, 2221, 2222, 2223

Interacciones catiónicas- π

Residuo	AA	Distancia	Compensar	Grupo del ligando	Átomo del ligando
119A	ARG	4.28	1.91	Aromático	2573, 2574, 2576, 2577, 2578, 2579

* **Distancia H-A:** Distancia entre el hidrógeno y los átomos aceptores

***Distancia D-A:** Distancia entre átomos donantes y aceptores

*[O3]: *sp* 3-oxígeno hibridado

*[Nam]: Nitrógeno de amida desprotonada

*[Npl]: Nitrógeno con hibridación *sp* 2 pero sin doble enlace (planar)

*[Ng+]: Nitrógeno de guanina

*[O2]: Oxígeno hibridado *sp* 2

*[Nar]: Nitrógeno aromático