



UNIVERSIDAD DE BUENOS AIRES  
FACULTAD DE CIENCIAS EXACTAS Y NATURALES

# Búsqueda racional de blancos terapéuticos para atacar al bacilo de la tuberculosis en la fase de latencia

Tesis presentada para optar al título de Magister en Biología Molecular  
Médica

**Autor:** Bioingeniero Ivan Matías Kruk

**Director:** Dr. Marcelo Martí

**Lugar de trabajo:** Grupo de Bioinformática Estructural, Departamento de  
Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad  
de Buenos Aires (UBA).

BUENOS AIRES, MARZO 2014

# Búsqueda racional de blancos terapéuticos para atacar al bacilo de la tuberculosis en la fase de latencia

## **RESUMEN:**

Se realizó un análisis de datos de sobre-expresión diferencial genómica del organismo *Mycobacterium Tuberculosis* (Mt), cepa H37Rv, en ensayos in-vitro en condiciones de estrés simulando el estado de latencia del bacilo en el huésped. Este análisis se combinó con información de diferentes bases de datos bioinformáticas de dominio público a través de una nueva herramienta en desarrollo denominada TuberQuery. Esta herramienta permitió identificar una serie de *Open Reading Frames* (ORFs) con una determinada homología de secuencia de *Protein Families* (PFAM) que tienen al menos una estructura molecular detallada en el *Protein Data Bank* (PDB). Posteriormente se consideró información metabólica y de esencialidad para clasificar los candidatos líderes obtenidos. Finalmente, se propone la continuación de este trabajo con investigaciones de *drogabilidad*, *off-targeting*, y *virtual screening* necesarios para estimar la potencialidad farmacológica de los candidatos líderes obtenidos. Se plantea el beneficio de utilizar herramientas bioinformáticas para la búsqueda de nuevos objetivos moleculares como una alternativa para optimizar recursos en el comienzo del proceso racional de descubrimiento de drogas.

# Rational approach to drug target discovery in latent *Mycobacterium tuberculosis*

## **ABSTRACT:**

A differential expression microarray analysis of stressed in-vitro *Mycobacterium Tuberculosis* (Mt), strain H37Rv, mimicking the microorganism's latent state in the host was performed. This analysis was combined with information from various publicly available bioinformatics databases using a new tool in development called TuberQuery. This tool allowed us to identify a number of Open Reading Frames (ORFs) with a certain Protein Family (PFAM) sequence homology, for which exists at least one molecular structure listed in the Protein Data Bank (PDB). Subsequently metabolic and essentiality information was considered to classify the lead compounds obtained. Finally, we propose the follow up of this work with further research on *drugability*, off-targeting and virtual screening in order to estimate the pharmacological potency of the lead candidates obtained. We state the benefit of utilizing bioinformatics tools to pursue the discovery of novel drug targets as one way to optimize the resources in the beginning stages of the rational drug discovery process.

## ÍNDICE DE CONTENIDO

1.	INTRODUCCIÓN .....	7
1.1	Búsqueda racional de nuevos objetivos moleculares para el desarrollo de drogas.....	7
1.2	Características de Mycobacterium tuberculosis y su patología en humanos ...	10
1.3	Tratamientos actuales y ensayos clínicos contra la Tuberculosis en humanos	13
1.4	Modelos y mecanismos moleculares de la TB latente .....	18
2.	MÉTODO.....	20
2.1	Propuesta para el desarrollo racional de drogas contra la TB latente .....	20
2.2	Criterios de selección de publicaciones de sobre-expresión genómica de Mt latente inducida experimentalmente .....	26
2.3	Criterios de selección de genes sobre-expresados .....	26
2.4	Criterios de selección de proteínas esenciales .....	27
3.	RESULTADOS .....	29
3.1	Publicaciones de expresión génica de Mt en estado de latencia inducida .....	29
3.2	Genes de Mt sobre-expresados en estado de latencia inducida .....	30
3.3	Proteínas de Mt esenciales sobre-expresadas en estado de latencia inducida..	36
4	DISCUSIÓN DE RESULTADOS .....	44
5	TRABAJO FUTURO.....	48
6	AGRADECIMIENTOS .....	50
7	REFERENCIAS .....	51

## ÍNDICE DE TABLAS, GRÁFICOS Y ANEXOS

### Tablas:

Tabla 1. Ensayos clínicos sobre TB publicados por el Instituto de Salud de los Estados Unidos (NIH). Total de 580 ensayos de TB listados al 09-Septiembre-2012 (“Clinicaltrials.gov Search Results” 2012).....	17
Tabla 2. Publicaciones analizadas de sobre-expresión génica de Mt en estado de latencia inducida in-vitro .....	29
Tabla 3. Tablas de sobre-expresión génica de Mt en estado de latencia in-vitro utilizadas en este trabajo .....	30
Tabla 4. Los genes identificados como sobre-expresados más de 200% en el mayor número de publicaciones analizadas. Para ver la lista de genes completa referirse al Anexo 1.....	33
Tabla 5. Tripletes UNIPROT-PFAM-PDB identificados a partir de cada tabla de sobre-expresión analizada en TuberQuery .....	34
Tabla 6. Tripletes UNIPROT-PFAM-PDB con score mayor a 1 según las tablas de sobre-expresión analizadas y TuberQuery. Para ver el listado completo referirse al Anexo 2.....	35
Tabla 7. Información estructural, funcional y de esencialidad de las proteínas del TuberQuery Job con score mayor a 1 .....	39
Tabla 8. Resumen de información estructural, funcional, metabólica, y de esencialidad de las proteínas del TuberQuery Job con score mayor o igual a 1 .....	42
Tabla 9. Resumen de candidatos a blanco molecular tras cada filtro de selección .....	42

Gráficos:

Gráfico 1. Flujo de trabajo de la base de datos TuberQuery ..... 23

Gráfico 2. Frecuencia de detección de sobre-expresión génica de Mt en estado de latencia inducida en número de publicaciones analizadas..... 31

Gráfico 3. Frecuencia de detección de sobre-expresión génica de Mt en estado de latencia inducida como porcentaje del número de publicaciones analizadas y limitándonos a mostrar una congruencia de al menos 2 publicaciones ..... 32

Gráfico 4. Captura de pantalla de TuberQuery para estructura PDB 3CIS asociada al gen locus MT2698..... 43

Anexos:

Anexo 1. Tabla completa de sobreexpresión SoloPapers - Tabla\_mayor2\_filtrada3

Anexo 2. Tabla completa de sobreexpresión TuberQuery- Todos\_con\_la\_base2

## **1. INTRODUCCIÓN**

### **1.1 Búsqueda racional de nuevos objetivos moleculares para el desarrollo de drogas**

El proceso de desarrollo de drogas goza históricamente de una muy baja eficiencia y un muy alto costo (Sauer 2005). Originalmente el descubrimiento de nuevas drogas se realizaba a través de prueba y error de sustancias químicas sobre cultivos celulares o animales y se anotaba y emparentaba la observación de la exposición al compuesto y el químico utilizado.

El paradigma actual de desarrollo de medicamentos, conocido como diseño racional de drogas, involucra la identificación de uno o un conjunto de blancos terapéuticos que al ser regulados por la acción de una droga o un cóctel de drogas específico produce el efecto terapéutico deseado. Este paradigma utiliza y muchas veces requiere de la bioinformática en casi todas sus etapas (Chen and Chen 2008a; Blundell et al. 2006; Scapin 2006; Nwaka and Ridley 2003). La identificación de proteínas blanco, desde la secuenciación y anotación funcional del genoma, requiere la aplicación y desarrollo de herramientas bioinformáticas asociadas al manejo de secuencias (Chen and Chen 2008b). La caracterización estructural de las proteínas blanco y sus posibles interacciones con posibles fármacos es otra rama donde la bioinformática estructural ha hecho grandes aportes, colaborando en el diseño racional de drogas (Tramontano 2006; Rester 2008; Agüero et al. 2008). Por su parte, el desarrollo de nuevas terapias combinadas, involucra el estudio de redes metabólicas o de señalización y la

evaluación de los posibles efectos de los cócteles de drogas, en donde la utilización de la biología de sistemas, una rama de la bioinformática, es moneda corriente (Anishetty, Pulimi, and Pennathur 2005). Estos tres enfoques mencionados, han sido, y continúan siendo utilizados, en general de forma aislada, para determinar posibles blancos moleculares (Ortí et al. 2009).

La definición de desarrollo de drogas por el método convencional y por el método de diseño racional sugiere en sí misma las ventajas del paradigma actual de desarrollo de drogas. La evolución del método científico para desarrollo de drogas hacia un método de diseño racional ha permitido la generación de drogas más selectivas, eficaces y eficientes (Mandal, Moudgil, and Mandal 2009). En la actualidad la mayoría, sino todas, las farmacéuticas y centros de investigación a nivel mundial utilizan un proceso de desarrollo de drogas racional que parte de una investigación informática inicial.

Dos herramientas fundamentales que permitieron avances espectaculares en el campo de la bioinformática aplicada al desarrollo de drogas en las últimas dos décadas fueron la secuenciación genómica (Dark 2013) y el microarreglo de expresión génica (Malone and Oliver 2011). Estas herramientas abrieron un abanico de oportunidades que aun están en pleno desarrollo. La aplicación de técnicas bioinformáticas en el estudio de nuevos compuestos con potencial valor terapéutico ha experimentado una enorme evolución en este período, abriendo la posibilidad de combinar diversas estrategias computacionales para el desarrollo de fármacos (Jorgensen 2009; Jorgensen 2004).



En gran parte esto ha estado relacionado con la creación y expansión constante de bases de datos bioinformáticas como el Protein Data Bank (PDB) que aumenta el número de estructuras disponibles periódicamente (Berman et al. 2000). Esto se ha traducido particularmente en una rápida evolución de los métodos de docking. Estos métodos permiten una exploración eficiente de mega-librerías de compuestos con el objetivo de determinar potenciales candidatos (compuestos *hits*) a interactuar con una determinada diana de interés farmacológico que emerge de estudios de genómica y proteómica (Taylor, Jewsbury, and Essex 2002). Hoy en día, el docking de librerías, algunas de carácter público como ZINC (Irwin and Shoichet 2005), que contiene cerca de 13 millones de compuestos, u otras de carácter privado en las multinacionales farmacéuticas, facilitan la generación de candidatos a fármacos basados en estructura. Métodos bioinformáticos como la dinámica molecular permiten explorar de forma teórica el modo de unión predicho por docking (Alonso, Bliznyuk, and Gready 2006), las metodologías de evaluación mediante cálculos de energía libre permiten determinar la afinidad de unión fármaco-diana (Vosmeer et al. 2014). Estas metodologías permiten predecir la capacidad de una diana de interactuar con compuestos tipo fármaco. La combinación de estas técnicas hace posible el diseño molecular *in-silico* por cuanto computan el impacto de modificaciones químicas estructurales y permiten al técnico-investigador filtrar y seleccionar candidatos que cumplan con atributos deseables para el desarrollo de nuevos fármacos o blancos moleculares (Walters, Ajay, and Murcko 1999).

La disponibilidad de dichas técnicas bioinformáticas es fundamental para seleccionar aquellas dianas que ofrecen mayores posibilidades de conseguir una modulación efectiva de la actividad biológica, e incidir así en una determinada patología (Seco, Luque, and Barril 2009). Estas herramientas están en general disponibles para su uso gratuito a través de internet y por ello son muy utilizadas en entornos académicos y en sectores industriales independientemente de su capacidad económica. Con el debido conocimiento son herramientas que acercan la brecha entre la necesidad social de nuevos fármacos y la realización de los mismos.

## **1.2 Características de *Mycobacterium tuberculosis* y su patología en humanos**

*Mycobacterium tuberculosis* (Mt) es una bacteria de la familia de Actinomicelados que causa una patología en humanos conocida como tuberculosis (TB). El organismo fue descubierto por Robert Koch en 1882. En vacas, el organismo causante de la tuberculosis bovina se denomina *Mycobacterium bovis* (Mb) descubierto en 1854 por Jean Antoine Villemin.

Mt es una bacteria no-motil de forma de bastón de 2-4  $\mu\text{m}$  de largo y 0.2-0.5  $\mu\text{m}$  de ancho. Es un parásito aeróbico facultativo usualmente de los macrófagos y se cree que por esta razón es frecuente encontrarlo en la zona de los glóbulos superiores de los pulmones. Su genoma posee aproximadamente 5,000 genes y ya fue secuenciado (Cole 1999; Cole et al. 1998). Tiene un tiempo de replicación relativamente lento de 15-20hs. Las bacterias Mt no son consideradas ni gram

positivas ni gram negativas, ya que poseen una química particular en su pared celular, principalmente compuesta por lípidos como el ácido micólico que no permite la entrada de la tinción gram positiva a pesar de contener peptidoglicano. Debido a su impermeabilidad a ciertos tintes se las conoce como *acid-fast bacteria* y al ser teñidas con carbol-fushin (teñido de Ziehl-Neelsen) adquieren un color rosado bajo el microscopio (magnificación 1000x). In vitro Mt se puede cultivar en medio Middlebrook, basado en agar, o Lowenstein-Jensen, basado en huevo, y toma 4-6 semanas obtener colonias visibles a simple vista.

El bacilo de Mt usualmente ataca a los pulmones, pero también puede hospedarse en el hígado, la columna vertebral, y el cerebro (“Centers for Disease Control and Prevention (CDC)” 2013).

Tras la exposición al bacilo de la TB, éste ingresa a través de los pulmones y se replica en macrófagos no activados dando lugar a una etapa de TB activa. Seguidamente el sistema inmune reconoce los agentes foráneos presentados por los macrófagos y forma granulomas en los cuales las bacterias son encontradas alrededor y dentro de linfocitos en un entorno de bajo porcentaje de oxígeno (Muttucumaru et al. 2004). Allí se detiene la replicación de la bacteria y comienza lo que se conoce como un estado de TB latente. Es aceptado mundialmente que el estado de latencia del Mt esta caracterizado por un bajo metabolismo de Mt (Park et al. 2003) y por un entorno de estrés caracterizado por una baja concentración de oxígeno (Sherman et al. 2001). Una persona con TB latente normalmente no transmite la enfermedad a otras personas. Bajo

ciertas condiciones, por ejemplo un sistema inmune reprimido, el bacilo de *Mt* logra volver a su estado de replicación y la patología a un estado de TB activa. En el caso de pacientes con TB activa, ésta es altamente contagiosa.

*Mt* es el organismo causante de la “plaga blanca” durante los siglos 17 y 18 en Europa, período durante el cual cerca del 100% de la población europea tenía TB y cerca del 25% de los adultos morían de TB (“Unión Internacional Contra La Tuberculosis Y Enfermedades Respiratorias”).

Actualmente la TB genera 8 millones de nuevas infecciones por año y causa más de 2 millones de muertes anuales (Park et al. 2003) principalmente en los países en desarrollo (Dye et al. 1999; “Global Tuberculosis Control -Epidemiology, Strategy, Financing” 2009). La Organización Mundial de la Salud (WHO) estima que un tercio de la población mundial, cerca de 1.9 billones de persona, son portadoras de *Mt* latente y pueden portar dicha infección asintomática por años o décadas (Park et al. 2003; Muttucumaru et al. 2004).

El aspecto determinante de la alta incidencia de tuberculosis en el mundo es la capacidad de supervivencia del organismo en un huésped humano sin causar efectos clínicos (Sherman et al. 2001), conocido como estado de latencia. En este micro-entorno celular el *Mt* se mantiene latente durante largos períodos, y actúa así como un reservorio de infección de gran relevancia para la subsistencia de dicha patología en humanos.

En la República Argentina, más de 12.000 personas enferman anualmente de Tuberculosis, y más de 1.000 mueren a causa de esta enfermedad (“Reporte Sobre Incidencia Y Mortalidad Por Tuberculosis En Argentina” 2008).

### **1.3 Tratamientos actuales y ensayos clínicos contra la Tuberculosis en humanos**

El diagnóstico de la presencia de Mt se realiza por un ensayo de piel conocido como Mantoux (aquellos vacunados con la BCG es posible que den positivo a este ensayo), o un ensayo por sangre (IGRAs) que no da falsos positivos en pacientes vacunados con la BCG. Para determinar si el paciente tiene TB activa es necesario tomar una tomografía de rayos X o una muestra de exudado.

En varios países del mundo es obligatoria la aplicación de la vacuna BCG (Bacille Calmette-Guerin) para intentar controlar la expansión de la enfermedad (“Centers for Disease Control and Prevention (CDC)” 2013). La vacuna de BCG se fabrica con cepas de Mb atenuadas y pre-cultivadas en medio Middlebrook 7H9. Fue utilizada en humanos por primera vez en 1921 y tras algunos eventos trágicos empezó a ser usada masivamente hacia 1948. Actualmente es efectiva en prevenir la TB de 0 a 80% en un período de hasta 15 años dependiendo de la región (genética de la población, exposición a otras infecciones, y diferencias en el medio ambiente) y el laboratorio involucrado en la preparación de la cepa (genética y medio de cultivo usados) (Venkataswamy et al. 2012; Fine 1995). La exposición a otras infecciones es particularmente importante por el fenómeno de *masking* causado por la exposición a micobacterias no tuberculinas (Palmer and

Long 1966). Países como los Estados Unidos optan por no aplicar la vacuna BCG ya que hay poca incidencia en ese país y en cambio priorizan identificar tempranamente a los pacientes con TB latente a través del ensayo de Mantoux y tratarlos para evitar así el reservorio del patógeno en el huésped humano en su territorio y el peligroso desarrollo de cepas multiresistentes (MDR).

El tratamiento contra la Mt depende del estado en que se encuentra la enfermedad, latente o clínicamente activa. El tratamiento para combatir la TB latente es el uso de medicamentos como Isoniazida (INH) durante 6 a 9 meses, Isoniazida y Rifapentina (RPT) durante 3 meses o Rifampicina (RIF) durante 4 meses (“Centers for Disease Control and Prevention (CDC)” 2013). En el caso de TB con estado clínico activo, hay 10 drogas actualmente aprobadas por la FDA. De estas drogas, la primera línea de defensa es Isoniazida, Rifampicina, Estambutol (EMB) y Pyrazinamida (PZA) durante 2 meses y una fase de continuación de 4 a 7 meses adicionales (“Centers for Disease Control and Prevention (CDC)” 2013).

Como se describe a continuación, el tratamiento de la infección tuberculina enfrenta hoy en día numerosos obstáculos (Nguyen and Pieters 2009). Actualmente los centros de salud presentan dificultades por el costo y la extensa duración de los tratamientos con las drogas actuales de primera línea (Isoniazida y Rifampicina) a lo que se suma la aparición cada vez más rápida de cepas MDR, que pueden ser tratadas con drogas de segunda generación con efectos secundarios severos como ceguera/sordera y de cepas extremadamente resistentes (XDR) para las que no existe tratamiento aún (Janin 2007). Por otro

lado existen interacciones negativas entre las drogas tuberculinas y las drogas utilizadas en pacientes con SIDA lo que presenta un problema severo en casos comunes de co-infección (Bonora and Di Perri 2008).

Finalmente un punto clave es que no existen drogas específicas que ataquen la Mt en el macrófago en la fase latente de la patología. Aproximadamente un 20% de las personas con Mt latente desarrollan la TB en su forma activa, y el porcentaje es aún mayor en personas inmunosuprimidas. Este enorme reservorio del patógeno hace muy difícil su erradicación. En este contexto, el desarrollo de nuevos fármacos es apremiante (“Global Tuberculosis Control -Epidemiology, Strategy, Financing” 2009). El desarrollo de nuevos fármacos antituberculosos también enfrenta numerosos obstáculos, siendo los más importantes el limitado conocimiento de la fisiología del bacilo especialmente en condiciones fisiológicamente relevantes como la de stress nitro-oxidativo, el desconocimiento de blancos moleculares para matar al bacilo en estado latente, y la falta de comprensión sobre por qué muchos fármacos prometedores *in-vitro* fallan en las etapas posteriores de desarrollo (*in-vivo*/clínico) (“Global Tuberculosis Control -Epidemiology, Strategy, Financing” 2009).

Las drogas actuales contra la tuberculosis son mayoritariamente moléculas de tamaño pequeño, de carácter bactericida general, con acción citotóxica que afectan principalmente bacterias en replicación. Su gran desventaja es que cuando son aplicadas como mono-droga o cuando su extenso tratamiento es interrumpido pueden generar bacterias resistentes, y no son suficientemente agresivas para erradicar la Mt latente.

Los mecanismos de acción de estos medicamentos se detallan a continuación:

- Rifampicina/Rifanpin: antibiótico de amplio espectro, obtenido sintéticamente de *Streptomyces mediterranei*. Inhibe la RNA polimerasa ADN dependiente. Actúa sobre bacterias intra y extracelulares. No inhibe la enzima de mamíferos.
- Isoniazida: es una pro-droga catalizada por KatG e hidrazina en bacterias e inhibe la producción de ácido micólico, esencial para la pared celular bacteriana. Actúa sobre bacterias intra y extracelulares.
- Estambutol: inhibe la arabinosyl transferasa que está involucrada en la síntesis de ácido micólico y spermidina. Actúa sobre bacterias intra y extracelulares.
- Pyrazinamida: sólo es activa contra Mt a un pH ligeramente ácido. Es una pro-droga activada por la enzima pyrazinamidasa en el bacilo Mt y convertido a ácido pyrazinoico (POA) que se cree se acumula en el citosol del bacilo, disrumpe la pared celular y la actividad de la sintasa FAS I (síntesis de lípidos).
- Estreptomina: es un antibiótico amino-glucósido que se une irreversiblemente e inhibe la subunidad 30S ribosomal bacteriana causando errores de unión al t-RNA en el complejo de iniciación generando proteínas sin la actividad deseada. Tiene efectos secundarios sobre el riñón y el oído.



Se cree que la subpoblación de células de Mt sobreviviente en estado de latencia tras la embestida del sistema inmune es la causante de la persistencia de le TB en humanos inclusive tras intensos regímenes de antibióticos y quimioterapia (Hampshire et al. 2004).

Estado de Ensayos Clínicos	Cantidad de Ensayos
Activos sin reclutar pacientes	73
Activos reclutando por invitación	6
Activos pero aun no reclutando	35
Activos y reclutando pacientes	165
Completados	281
Cancelados	20
Total:	580

Tabla 1. Ensayos clínicos sobre TB publicados por el Instituto de Salud de los Estados Unidos (NIH). Total de 580 ensayos de TB listados al 09-Septiembre-2012 (“Clinicaltrials.gov Search Results” 2012)

La gran mayoría de estos ensayos clínicos son estudios de nuevos medicamentos contra la tuberculosis. De los 580 ensayos clínicos publicados, solo 19 han publicado los resultados del mismo en el sitio de ensayos clínicos del NIH consultado.

La cantidad de ensayos clínicos publicados en el NIH a la fecha citada con referencia a TB son pocos si se comparan con otras patologías como HIV (5514), cáncer de mamas (4929), influenza (1784), Hepatitis B (1391) o neumonía (1231).

#### 1.4 Modelos y mecanismos moleculares de la TB latente

En 1967 Wayne et al. desarrolló una condición de cultivo in-vitro de Mt persistente pero no replicativo usando bajas concentraciones de oxígeno (Wayne and Diaz 1967). En este estudio se mostró que si la Mt in-vitro es expuesta bruscamente a un estado anaeróbico, no sobrevive, pero si es llevada por un estado de depleción gradual de oxígeno llamado *non-replicative stage 1* o NRP1 la Mt logra adaptarse y sobrevivir a las nuevas condiciones del microentorno. Con lo cual se considera que el paso por un estado gradual de baja concentración de oxígeno NRP1 es fundamental para su adaptación y supervivencia en condiciones de hipoxia in-vitro. Durante NRP1 la densidad óptica del cultivo de Mt sigue creciendo debido no a replicación celular sino a un incremento de tamaño de célula principalmente debido a un engrosamiento de la pared celular característico de un estado de hipoxia. Durante esta etapa no hay síntesis de DNA, hay menos síntesis de RNA, y hay una actividad enzimática modulada. Durante el siguiente período, denominado *non-replicative stage 2* o NRP2 la densidad óptica se mantiene constante, cambia el transcriptoma respecto a NRP1 y el cultivo adquiere resistencia al antibiótico Isoniazida y sensibilidad al antibiótico Metronidazol (pro-droga con actividad bactericida conocida contra bacterias anaeróbicas activada por reducción química vía ferredoxinas). Durante NRP2 la Mt puede permanecer latente durante largos períodos de tiempo y reactivarse y proliferar al encontrarse en condiciones de oxígeno normales nuevamente (Muttucumaru et al. 2004).

Hay una diferencia importante entre el estado de latencia de la Mt y un estado de estrés nitro-oxidativo generalizado. En el primero de los casos se activa un subconjunto de genes que se conocen como el *dormancy regulon* e involucra 48 genes. En el segundo caso se activa un número mucho mayor de genes, más de 400, que están relacionados a una respuesta genérica a un estrés extracelular como fue publicado tras estudios del transcriptoma de la Mt ante diferentes condiciones de experimentación (Voskuil et al. 2003).

Uno de los genes inducido por condiciones de hipoxia y mas estudiado es el gen Rv3133/DosR denominado *dormancy survival regulon*, un factor de transcripción que modula la respuesta del regulón dosR cuando el Mt se encuentra en estado de estrés nitro-oxidativo (Park et al. 2003).

Otro factor que se cree podría desencadenar o inducir el estado de latencia del Mt es una elevada concentración de óxido nítrico (NO). El NO es generado por macrófagos activados debido a su característica bactericida. Evidencia de esto son los modelos de infección progresiva en ratones con deficiencia de NO sintetasa (NOS) o en ratones *wild type* tratados con un inhibidor de NOS (iNOS) (Voskuil et al. 2003). Adicionalmente el NO y las especies intermedias de nitrógeno (RNI) son agentes con acción antimicótica y antibiótica y juegan un papel intracelular importante como moléculas de transducción de señales (Ohno et al. 2003).

Aún se conoce poco del entorno bioquímico del macrófago infectado con Mt, es decir la concentración de iones, radicales, lípidos y metabolitos. Se sabe que el

Mt al ser inhalado en humanos es fagocitado por el macrófago alveolar, cuya maduración el Mt es capaz de detener para poder sobrevivir y proliferar. No obstante la defensa inmunológica celular-adaptativa, mediante la citoquina interferón gama, permite al macrófago alcanzar su madurez y detener la propagación de Mt. Aun así un porcentaje del Mt sobrevive y queda en un estado de latencia no replicativa indeterminado hasta su nueva reactivación (Schnappinger et al. 2003).

## **2. MÉTODO**

### **2.1 Propuesta para el desarrollo racional de drogas contra la TB latente**

Planteamos la utilización de la información científica disponible en publicaciones y bases de datos biológicas, y la capacidad de procesamiento y modelado de estos datos por computadora, para lograr focalizar los esfuerzos y maximizar los resultados de desarrollo de nuevos fármacos para nuevos blancos moleculares.

Planteamos la utilidad de generar una lista consenso de sobre-expresión de proteínas de Mt cuando éste se encuentra en un estado de estrés in-vitro similar al estado de latencia que enfrenta en el huésped. Para las proteínas derivadas de dicha lista se analiza la posibilidad de ser un objetivo drogable. En este trabajo por drogable nos referimos principalmente al índice de drogabilidad, es decir la propiedad de la proteína blanco de poseer un bolsillo que permita la unión de una pequeña molécula tipo droga que cumpla con la regla de Lipinski.

Posteriormente se analiza la esencialidad y rol metabólico del target. El termino esencialidad se refiere a la dependencia del organismo Mt de un gen en particular para su crecimiento y/o supervivencia, y por consiguiente si ese gen es inhibido resulta en una efecto bacteriostático o bactericida (Agüero et al. 2008).

Planteamos que esta metodología nos permitirá filtrar y seleccionar los objetivos moleculares con mayores probabilidades de éxito como nuevos blancos moleculares contra la TB latente. Estos compuestos líderes deberán ser sometidos a un posterior análisis estructural *in-silico* más profundo para estudiar la factibilidad de interacción droga-proteína, y su posterior prueba de concepto in-vivo y desarrollo farmacológico.

Para llevar adelante los objetivos generales propuestos, se desarrollará un esquema de puntuación y selección de posibles blancos, que comenzará con el estudio a nivel secuencia del genoma completo de Mt cruzando información de las diferentes bases de datos, genómicas, de expresión, estructurales y de vías metabólicas para la obtención de un subconjunto ordenado de proteínas de interés, asociadas principalmente a las condiciones experimentales de estrés inducido por hipoxia.

Para identificar aquellos genes sobre-expresados que sean de interés farmacéutico es necesario considerar la *drogabilidad* de los potenciales blancos moleculares. Para ello es de gran utilidad conocer la estructura molecular de las proteínas sobre-expresadas en estado de latencia inducida. La cristalización y obtención de estas estructuras es muy laborioso y difícil de automatizar y

producir en escala. Un gran número de los genes predichos en Mt codifican proteínas con funciones y estructuras desconocidas (Cole et al. 1998). Por lo tanto nuestro enfoque racional de descubrimiento de nuevos blancos moleculares por métodos bioinformáticos involucra identificar estructuras moleculares conocidas que contengan alguna similitud con las secuencias conservadas de los ORFs de Mt identificados como sobre-expresados en los experimentos de interés.

Parte de esta estrategia se desarrolla utilizando una nueva base de datos, TuberQuery (“Tuberquery” 2014), generada por el grupo de investigación de Martí et al. en paralelo a esta tesina y que está a disposición de los interesados y que creemos puede servir como una fuente de información útil para otros grupos de investigación trabajando en el descubrimiento de drogas contra la TB. Asimismo creemos que TuberQuery sirve como modelo de base de datos para enfrentar la difícil tarea de desarrollo de nuevas drogas en otras patologías.

TuberQuery utiliza información depositada en diferentes bases de datos biológicas como Tuberculosis Database (TBDB), TubercuList, Uniprot, Gene-Ontology (GO), Protein Family (PFAM), Protein Databank (PDB), RefSeq-NCBI, NCBI Gene Expression Omnibus (GEO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Lanza pockets, Modbase, Catalytic Site Atlas (CSA), PropKa, y Drugbank.

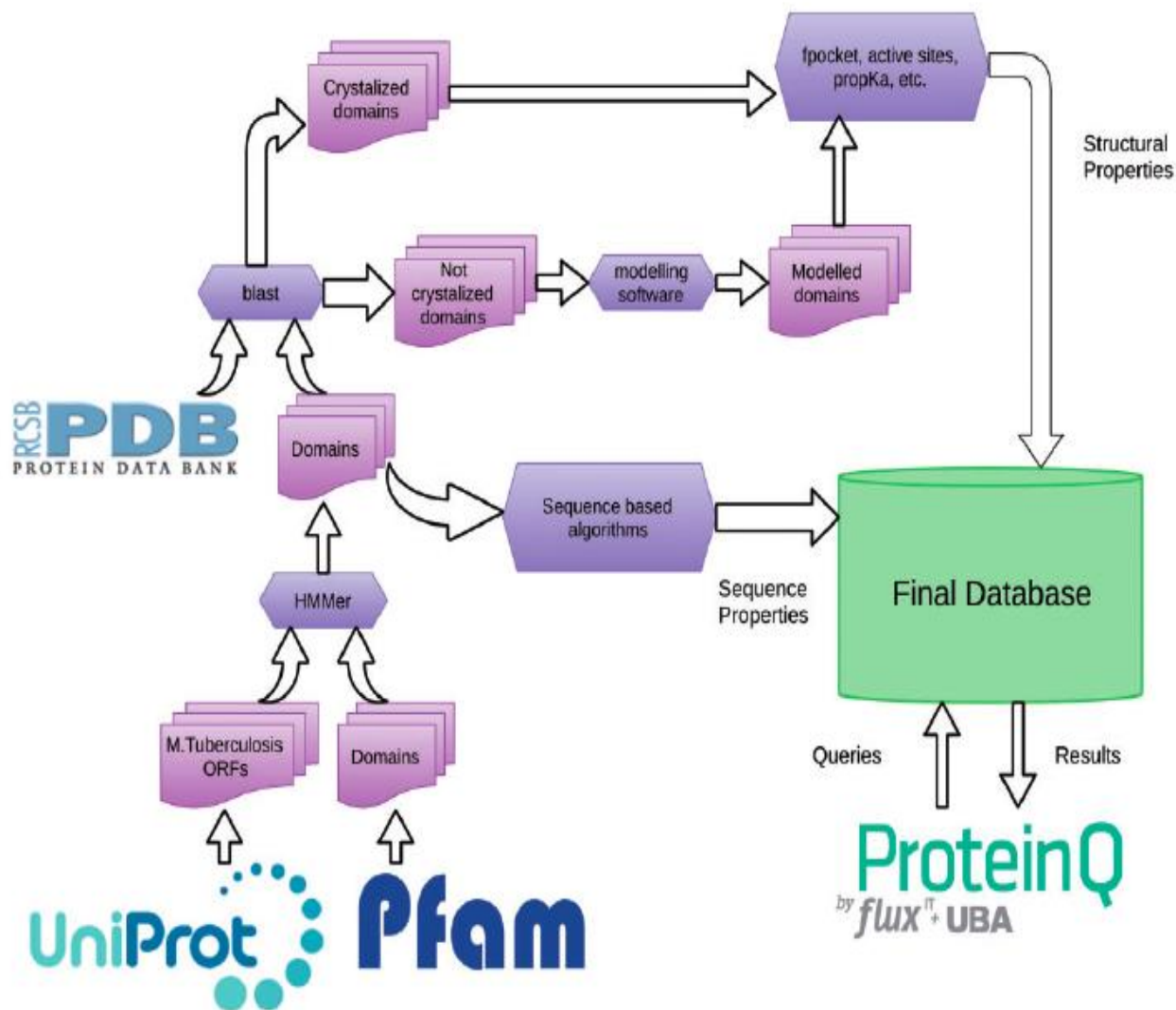


Figure 1. A schematic representation of the TuberQ Pipeline.

Gráfico 1. Flujo de trabajo de la base de datos TuberQuery

La base de datos TuberQuery se alimenta de los ORFs de Mt anotados en la base de datos TubercuList y los dominios de las familias de proteínas anotadas en la base de datos Pfam (más de 13,000 entradas entre Pfam-A y Pfam-B). Mediante un algoritmo probabilístico llamado Modelos Ocultos de Markov TuberQuery determina, para cada ORF en el genoma de Mt, a que familias Pfam pertenece dicho ORF mediante el alineamiento de los diferentes dominios proteicos de las

familias Pfam y la secuencia proteica putativa derivada del ORF de Mt. Las secuencias proteicas de estos dominios encontrados en las proteínas putativas de Mt son luego utilizadas para realizar un psi-blast sobre todas las secuencias proteicas anotadas en la base de datos estructural PDB (más de 77,000 entradas en su base de datos). Esto permite obtener un triplete Uniprot-PFAM-PDB con una cierta identidad y cobertura sobre los dominios analizados. De no existir una estructura cristalizada en el PDB para un determinado dominio Pfam identificado en el paso anterior, se utiliza diferentes programas de modelado primero para determinar la capacidad de modelado de la secuencia proteica derivada del ORF y de ser posible para modelar su estructura.

Consideramos como significativos resultados HMMER *e-value*  $< 1 \times 10^{-7}$  y resultados BLAST *e-value*  $< 1 \times 10^{-7}$ . Los valores e mencionados son la cantidad de *hits* que se espera tendrían una coincidencia igual o mejor que el resultado encontrado por azar dada la cantidad de entradas en la base de datos. Por lo tanto cuanto menor a 1 sea este valor, mayor es la significancia estadística del resultado encontrado.

A este universo de estructuras proteicas identificadas y provenientes del PDB o del modelado de estructuras de TuberQuery, se le computa el índice de drogabilidad de todos los pockets encontrados en su estructura utilizando el algoritmo de fpocket. Finalmente a cada estructura se le adiciona una serie de datos descriptivos de otras bases de datos como propKa (valores de pKa de grupos ionizables en la proteína), residuos y sitios activos anotados en la base de datos CSA para dicha proteína, etc. Todos estos datos junto con la



representación gráfica de las estructuras encontradas o generadas son mostrados como resultado de la búsqueda inicial en TuberQuery.

La base de datos TuberQuery puede ser consultada de dos formas. Una de ellas es haciendo una consulta con un determinado Rv, Uniprot, Pfam, o PDB ID de interés. En este caso TuberQuery devuelve todos los ID asociados mediante el método de asociación detallado anteriormente, junto con la información de HMMER e-value, el Blast e-value, la identidad y la cobertura del Blast, el índice de drogabilidad de los pockets detectados, y un enlace desde donde se accede a la representación gráfica de la estructura y sus pockets así como información adicional de interés farmacológico proveniente de las distintas bases de datos que alimentan TuberQuery.

El otro método de consulta, y el cual fue utilizado en este trabajo involucra la determinación de valores para cada uno de los parámetros de selección que ejecuta la base de datos TuberQuery. Esto no devuelve información referente a un blanco en particular con sus IDs asociados, sino todos los tripletes Uniprot-PFAM-PDB que cumplen con las condiciones especificadas introducidas en la consulta.

## **2.2 Criterios de selección de publicaciones de sobre-expresión genómica de Mt latente inducida experimentalmente**

Se realizó una selección de publicaciones de revistas arbitradas científicamente cuyo contenido lista niveles de expresión genómica (respuesta transcripcional) de Mt en estado de estrés. Se utilizó el siguiente criterio de selección:

Publicaciones listadas en PUBMED en el período [2000-2012] con información disponible de los experimentos de microarreglos de sobreexpresión génica en Mt en estado de estrés, particularmente inducido por hipoxia. Como base se usó el listado de publicaciones referenciadas en el PICT 2010-2805 (M. Martí and Turjanski 2010). Teniendo en cuenta ese conjunto global de publicaciones científicas con información de expresión génica en experimentos de Mt con latencia inducida como punto de partida y considerando las restricciones originadas por la creación en paralelo de la base de datos TuberQuery, se decidió para esta etapa preliminar utilizar un subconjunto reducido de publicaciones con el objetivo de lograr una demostración de concepto en los tiempos disponibles para esta tesina.

## **2.3 Criterios de selección de genes sobre-expresados**

Se combinó la información de las listas de expresión de las publicaciones seleccionadas en el punto anterior para generar un subconjunto de genes de TB con sobre-expresión en estado de estrés. Para ello se utilizó el siguiente criterio de selección:

Genes expresados al menos un 200% en condiciones experimentales de latencia inducida en comparación con condiciones de crecimiento exponencial de Mt en cultivo.

Cada vez que un ORF aparecía sobre-expresado en una publicación dentro del subconjunto de publicaciones analizadas, se le asignaba un punto al ORF.

El objetivo del método de *data mining* utilizado anteriormente es destacar los genes que fueron encontrados sobre-expresados por un consenso de experimentos de TB en estado de hipoxia inducida realizado por diversos grupos de investigación. En esta selección no se consideró información estructural derivada de los ORFs.

Para encontrar genes de Mt sobre-expresados en condiciones de hipoxia inducida y que cuentan con información estructural relacionada, se procedió a la utilización de la base de datos TuberQuery. En este caso, se realizó una búsqueda considerando las mismas tablas de microarreglos, los mismos criterios de sobre-expresión, y valores de identidad de 0.6 y cobertura 0.8 para el psi-blast.

#### **2.4 Criterios de selección de proteínas esenciales**

La información de esencialidad utilizada para este trabajo fue obtenida a través de estudios que utilizan el método de mutagénesis por hibridación con transposones (TraSH) (Sasseti, Boyd, and Rubin 2003; Griffin et al. 2011).

Existen otros métodos disponibles y utilizados en Mt para la identificación de genes esenciales como es el análisis estequiométrico de balance de flujo (FBA) (Draghici et al. 2007; Raman, Rajagopalan, and Chandra 2005; Jamshidi and Palsson 2007; Beste et al. 2007), la determinación de proteínas *chokepoint* (Hasan et al. 2006), o la determinación por consenso del común de genes expresados bajo diferentes condiciones de crecimiento (Gao et al. 2005).

Para la clasificación de vías metabólicas se utilizaron los modelos GSMN-TB (Beste et al. 2007) y Mt INJ661 (Jamshidi and Palsson 2007). Entre estos dos modelos cubren 20% de los genes de Mt. También se utilizó la información del *mycolic acid pathway*, MAP (Raman, Rajagopalan, and Chandra 2005).

Para las proteínas que no pudieron ser mapeadas con los modelos indicados anteriormente se utilizó la información de las bases de datos KEGG (“KEGG Pathway Database” 2013; “The Kyoto Encyclopedia of Genes and Genomes—KEGG” 2000), Biocyc (“Biocyc Database Collection” 2013), GO, PFAM, Uniprot y Tuberculist para asistir en la determinación de esencialidad/actividad metabólica.

Estas bases de datos también se utilizaron para determinar si las proteínas en cuestión tienen grupos Hemos, *clusters* Fe-S, centros de cobre como cofactores, o residuos de tirosina o cisteína en el grupo activo. Estas características convierten a los potenciales blancos moleculares en moléculas susceptibles a ser afectadas por RNOS.

### 3. RESULTADOS

#### 3.1 Publicaciones de expresión génica de Mt en estado de latencia inducida

Búsquedas realizadas el 27-Agosto-2012.

Base de datos Pubmed con más de 20 millones de publicaciones a la fecha.

“Mycobacterium tuberculosis”: 42,179 hits

“Mycobacterium tuberculosis” AND “microarray”: 249 hits

“Mycobacterium tuberculosis” AND “microarray” AND “latent”: 18 hits

Listado de publicaciones accesible en (“Pubmed Search Results 1” 2012)

“Tuberculosis”: 198,917 hits

“Tuberculosis” “AND “microarray”: 292 hits

“Tuberculosis” “AND “microarray” AND “latent”: 18 hits

Listado de publicaciones accesible en (“Pubmed Search Results 2” 2012)

Al momento de la realización de este trabajo las tablas de sobre-expresión analizadas y cargadas en la base de datos de TuberQuery fueron las siguientes:

Id	Publicación	Condición Experimental
2a	Voskuil 2003 (O2)	Hipoxia In-vitro
3	Ohno 2003	NO In-vitro
5	Park 2003	Hipoxia In-vitro
6b	Voskuil 2004: (12 días NRP)	Hipoxia In-vitro

Tabla 2. Publicaciones analizadas de sobre-expresión génica de Mt en estado de latencia inducida in-vitro

En las tablas de sobre-expresión génica listadas se identificaron un total de 3944 ORFs únicos. Los ORFs se anotaron usando el código Rv y el locus genómico del organismo Mt cepa H37Rv (Cole et al, 1998). Esta misma anotación es utilizada por la plataforma TubercuList del Pasteur Francia, la plataforma KEGG de la Universidad de Kyoto Japón (Organismo T00015) y el NCBI del Instituto Nacional de Salud de Estados Unidos (genoma uid 57777).

### 3.2 Genes de Mt sobre-expresados en estado de latencia inducida

Al aplicar el criterio de selección de genes sobre-expresados al menos un 200% en las citas analizadas se obtuvieron 355 *hits* y 244 ORFs únicos sobre-expresados.

Id	Publicación	Hits Sobre-Expresión > 200%
2a	Voskuil 2003 (O2)	45
3	Ohno 2003	36
5	Park 2003	160
6b	Voskuil 2004: (12 días NRP)	114
Total:		355

Tabla 3. Tablas de sobre-expresión génica de Mt en estado de latencia in-vitro utilizadas en este trabajo

El siguiente gráfico muestra la frecuencia de identificación de los 244 genes en las 4 publicaciones analizadas.

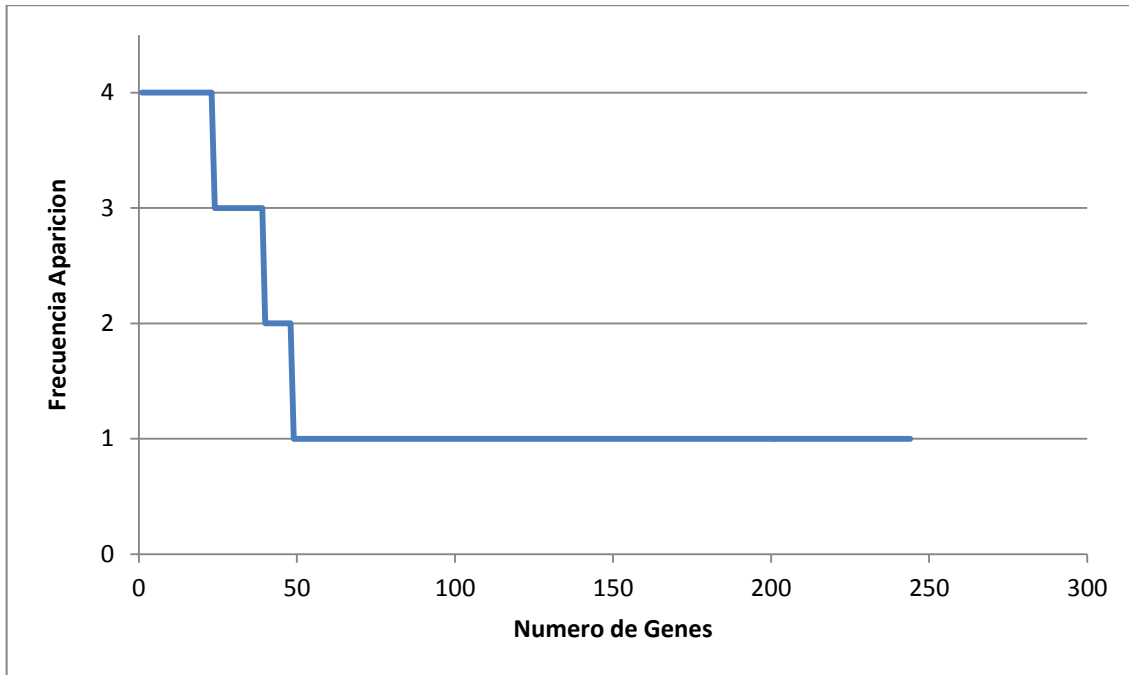


Gráfico 2. Frecuencia de detección de sobre-expresión génica de Mt en estado de latencia inducida en número de publicaciones analizadas

Considerando solamente aquellos genes que fueron reportados sobre-expresados en al menos dos publicaciones analizadas obtuvimos el siguiente gráfico de número de genes vs frecuencia de aparición en publicaciones, siendo este último valor 100% si se obtenía el gen en cuestión reportado como sobre-expresado al menos un 200% en 4/4 tablas analizadas en esta etapa.

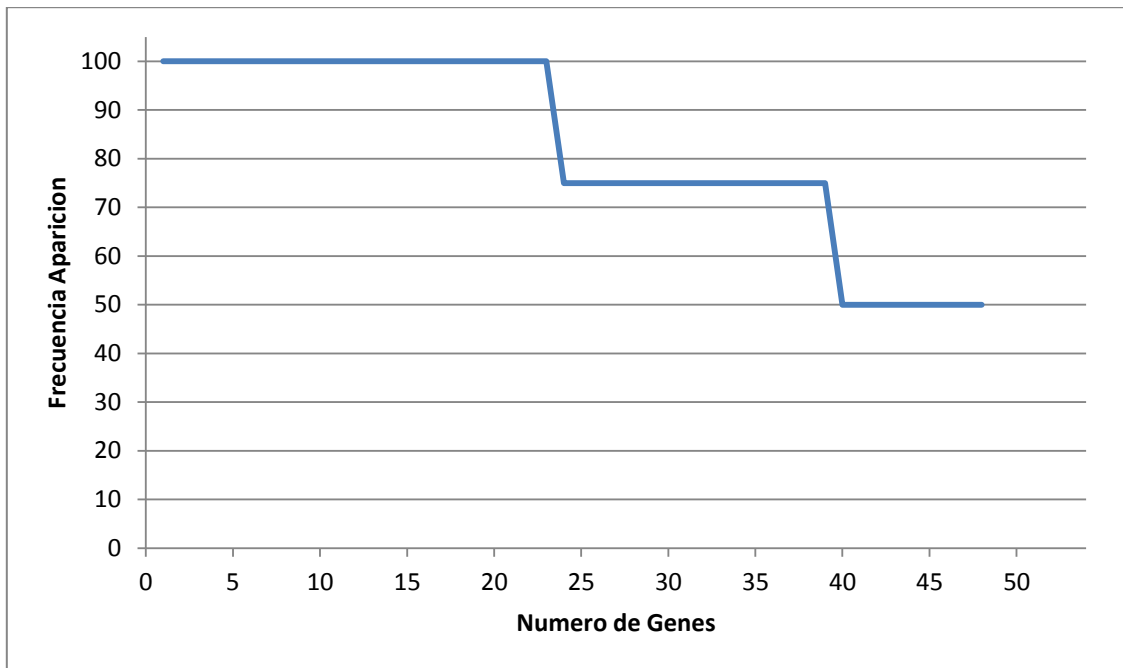


Gráfico 3. Frecuencia de detección de sobre-expresión génica de Mt en estado de latencia inducida como porcentaje del número de publicaciones analizadas y limitándonos a mostrar una congruencia de al menos 2 publicaciones

Para acotar la lista de candidatos a mostrar en la siguiente tabla acotamos el listado a aquellos genes detectados sobre-expresados al menos un 200% en las 4 publicaciones analizadas en esta etapa. Es de destacar que todos los genes identificados de esta manera corresponden al regulón DosR como se muestra a continuación.



Publicaciones	Locus	Proteína	Función
4/4	Rv0572C	experimental	Regulón DosR. Función desconocida
4/4	Rv1736C	narX	Regulón DosR. Probable narX, reductasa de nitrato
4/4	Rv1737C	narK2	Regulón DosR. Probable narK2, proteína integral de membrana transportadora de nitrato
4/4	Rv1997	ctpF	Regulón DosR. Proteína integral de membrana ATPasa transportadora de cationes metálicos
4/4	Rv2003C	experimental	Regulón DosR. Función desconocida
4/4	Rv2004C	experimental	Regulón DosR. Función desconocida. Contiene el motivo A del sitio de unión PS00017 ATP/GTP
4/4	Rv2005C	experimental	Regulón DosR. Proteína de stress universal
4/4	Rv2007C	fdxA	Regulón DosR. Ferredoxina, intermediario en metabolismo y respiración
4/4	Rv2028C	experimental	Regulón DosR. Proteína de stress universal
4/4	Rv2030C	experimental	Regulón DosR. Función desconocida
4/4	Rv2031C	hspX, alpha-crystallin gene, acr	Regulón DosR. Propuesta como encargada de mantener viabilidad a largo plazo en condiciones de hipoxia. Regulado por sistema regulador de 2 componentes: DEVR-Rv3133c / DEVS-Rv3132c
4/4	Rv2032	acg	Regulón DosR. Superfamilia de nitroreductasa
4/4	Rv2624C	experimental	Regulón DosR. Proteína de stress universal
4/4	Rv2625C	experimental	Regulón DosR. Probable proteína de transmembrana conservada rica en alanina y leucina
4/4	Rv2627C	experimental	Regulón DosR. Función desconocida
4/4	Rv2628	experimental	Regulón DosR. Función desconocida
4/4	Rv2629	experimental	Regulón DosR. Función desconocida
4/4	Rv3127	experimental	Regulón DosR. Función desconocida. Reportada como esencial para crecimiento in-vitro
4/4	Rv3129	experimental	Regulón DosR. Función desconocida
4/4	Rv3130C	tgs1	Regulón DosR. Sintetasa de triglicéridos
4/4	Rv3131	experimental	Regulón DosR. Función desconocida. Reportada como esencial para crecimiento in-vitro
4/4	Rv3132C	devS, dosS	Regulón DosR. Kinasa de 2 componentes unida a membrana sensor de histidina. Contiene el dominio GAF que une grupo hemo
4/4	Rv3133C	devR, dosR	Regulón DosR. Regulador de transcripción de 2 componentes

Tabla 4. Los genes identificados como sobre-expresados más de 200% en el mayor número de publicaciones analizadas. Para ver la lista de genes completa referirse al Anexo 1

Los resultados mostrados en la tabla anterior corresponderían a aquellos genes que se reportaron sobre-expresados al menos un 200% la mayor cantidad de veces en las citas analizadas. Esta selección de potenciales blancos no contempla ningún análisis sobre su estructura.

Para incluir datos de estructura en el análisis y selección de potenciales blancos moleculares se utilizó la base de datos TuberQuery sin filtro de datos de sobre-expresión y se obtuvieron: 1144 *hits* UNIPROT-PFAM-PDB, donde hay 451 ORFs únicos representados. Este es un grupo con relación muchos a muchos, es decir que un Uniprot puede pertenecer a más de una familia Pfam y esta familia puede tener más de una estructura PDB asociada y viceversa, es decir, varios Uniprot pueden pertenecer a la misma familia PFAM y varias familias Pfam pueden compartir la misma estructura PDB. Al incorporar el filtro de tablas de sobre-expresión en estado de latencia inducida experimentalmente y requerir a TuberQuery que seleccione solo aquellos ORFs sobre-expresados al menos un 200% obtuvimos los siguientes resultados:

Id	Publicación	Hits UNIPROT-PFAM-PDB
2a	Voskuil 2003 (O2)	7
3	Ohno 2003	2
5	Park 2003	42
6b	Voskuil 2004: (12 días NRP)	18
Total:		69

Tabla 5. Tripletes UNIPROT-PFAM-PDB identificados a partir de cada tabla de sobre-expresión analizada en TuberQuery

En la siguiente tabla destacamos los tripletes encontrados con score mayor a 1.

Uniprot Name	Locus Name (ORF)	Uniprot Id	Pfam Id	PDB Id	Blast Coverage	Blast Identity	Score
<a href="#">Transcriptional regulatory protein devR (dosR)</a>	Rv3133c	<a href="#">Q7D625</a>	<a href="#">PF00072</a>	<a href="#">3C3W</a>	1	0.85	4
<a href="#">Ferredoxin (fdxA)</a>	Rv2007c	<a href="#">Q10839</a>	<a href="#">PF00037</a>	<a href="#">2V2K</a>	1	0.68	4
<a href="#">Hypoxic response protein 1 (hrp1)</a>	Rv2626c	<a href="#">Q7D6V4</a>	<a href="#">PF00571</a>	<a href="#">1XKF</a>	1	1	3
<a href="#">Uncharacterized protein Rv0569/MT0595</a>	Rv0569	<a href="#">Q7D9M3</a>	<a href="#">PF08940</a>	<a href="#">2A7Y</a>	1	0.65	3
<a href="#">Hypoxic response protein 1 (hrp1)</a>	Rv2626c	<a href="#">Q7D6V4</a>	<a href="#">PF00571</a>	<a href="#">1Y5H</a>	0.98	0.94	3
<a href="#">Universal stress protein Rv2623/MT2698</a>	Rv2623	<a href="#">Q7D6V7</a>	<a href="#">PF00582</a>	<a href="#">3CIS</a>	0.86	1	3
<a href="#">Universal stress protein Rv2623/MT2698</a>	Rv2623	<a href="#">Q7D6V7</a>	<a href="#">PF00582</a>	<a href="#">2JAX</a>	0.86	0.98	3
<a href="#">Probable bacterioferritin (BfrB)</a>	Rv3841	<a href="#">Q7D4R7</a>	<a href="#">PF00210</a>	<a href="#">3QD8</a>	1	1	2

Tabla 6. Tripletes UNIPROT-PFAM-PDB con score mayor a 1 según las tablas de sobre-expresión analizadas y TuberQuery. Para ver el listado completo referirse al Anexo 2

El score representa una suma aritmética de la cantidad de publicaciones consideradas para este trabajo que identificaron como sobre-expresado al ORF en cuestión al menos un 200%. El resultado obtenido de TuberQuery fue una lista de 69 tripletes UNIPROT-PFAM-PDB con 52 tripletes únicos y 23 proteínas de Mt únicas. De este grupo destacamos que sólo 6/23 proteínas están codificadas en el regulón dosR, 5/23 proteínas no tienen información clara de su función reportada en Uniprot y 7/23 proteínas son reconocidas como blancos moleculares farmacológicos de alta confianza en Uniprot.

### 3.3 Proteínas de Mt esenciales sobre-expresadas en estado de latencia inducida

A continuación detallamos la información recopilada de función, metabolismo, esencialidad y características estructurales básicas para los genes identificados como potenciales blancos terapéuticos.

Analizando las proteínas con mayor consenso de sobreexpresión (listadas en la tabla anterior) podemos hacer las siguientes afirmaciones:

-Las seis proteínas identificadas como potenciales blancos moleculares según los criterios previamente detallados pertenecen al regulón dosR.

-Solo una de las seis proteínas (MT2698) es reconocida como esencial para la supervivencia de Mt in vivo por el método de TraSH.

-Cabe aclarar que en esta tesina no se han eliminado potenciales blancos moleculares por similitud de homología de blancos en el huésped (*off-targeting*).

Los detalles de cada proteína se muestran en la siguiente tabla.

Proteína:	devR/dosR
Gen Locus:	Rv3133c
Función:	Expresión esencial para la regulación de más de 50 genes específicos de la respuesta del Mt a condiciones de hipoxia
Esencialidad:	No esencial por mutagénesis transposomal Himar1 en H37Rv
Metaboloma:	No hay data en estudio de FBA citado
Estructura:	PDB 3C3W. Cristalografía de rayos X con resolución de 2.2Å (Univ. de Washington, Seattle USA) sugiere re-arreglo en alpha hélice con activación de posible dímero por fosforilación

Cant. aa en pdb:	225
Grupo Hemo:	no posee
Ligandos:	Liga SO4
Cu en cofactor:	no posee
Pfam data:	GerE (PF00196), Response_reg (PF00072)
Cys/Tyr sitio activo:	no hay entrada en CSA

Proteína:	ferrodoxin (FdxC)
Gen Locus:	Rv2007c
Función:	Provee poder reductor al citocromo P450 (CYP), catalizador de la oxidación de sustancias orgánicas como lípidos y sustancias xenobióticas
Esencialidad:	No esencial por mutagénesis transposomal Himar1 en H37Rv
Metaboloma:	No hay data en estudio de FBA citado
Estructura:	PDB 2V2K. FdxA es un ortólogo de Mycobacterium smegmatis y tiene dos clusters de hierro-azufre. Cristalografía de rayos X con resolución de 1.6Å (Univ. de Milano, Italia)
Cant. aa en pdb:	105
Grupo Hemo:	no posee
Ligandos:	Liga FE3-S4 CLUSTER (CLUSTER Fe-S) e ion acetato
Cu en cofactor:	no posee
Pfam data:	Fer4 (PF00037): 4Fe-4S binding domain, 4Fe-4S dicluster domain
Cys/Tyr sitio activo:	no hay entrada en CSA

Proteína:	hypoxic response protein 1 (HRP1)
Gen Locus:	Rv2626c
Función:	Altamente sobre-expresada en condiciones de estrés oxidativo. Su función no es aun del todo clara pero se sabe que es secretada al exterior (no es acumulada en la pared de la bacteria) y se cree que contribuye a la modulación del sistema inmune en el huésped. Se utiliza como antígeno para detectar estado de latencia en pacientes infectados.
Esencialidad:	No esencial por mutagénesis transposomal Himar1 en H37Rv
Metaboloma:	No hay data en estudio de FBA citado
Estructura:	PDB 1XKF. Proteína de la familia de "cystathionine-beta-synthase-domain-only", tiene 2 puentes disulfuro, y no parece interactuar con AMP. Cristalografía de rayos X con resolución de 1.9Å (Univ. de Auckland, Nueva Zelanda)
Cant. aa en pdb:	266
Grupo Hemo:	no posee
Ligandos:	Liga ion zinc
Cu en cofactor:	no posee
Pfam data:	CBS (PF00571)
Cys/Tyr sitio activo:	no hay entrada en CSA

Proteína:	MT0595
Gen Locus:	Rv0569
Función:	Tiene funcionalidad biológica desconocida pero resultados de microarreglos asocian su sobreexpresión con condiciones de falta de nutrientes y presencia de <i>heat shock proteins</i> . Posible unión a ADN para modulación de expresión de Rv0570 durante estado de hipoxia.
Esencialidad:	No esencial por mutagénesis transposomal Himar1 en H37Rv
Metaboloma:	No hay data en estudio de FBA citado
Estructura:	PDB 2 <sup>a</sup> 7Y. Caracterizada por NMR y dicroísmo circular surgieron que esta proteína hipotética tiene una estructura molecular novedosa (Pacific Northwest National Laboratory, Richland USA)
Cant. aa en pdb:	88
Grupo Hemo:	desconocido
Ligandos:	desconocido
Cu en cofactor:	desconocido
Pfam data:	dominio sin acción conocida (PF08940/DUF1918)
Cys/Tyr sitio activo:	no hay entrada en CSA

Proteína:	MT2698
Gen Locus:	Rv2623
Función:	Proteína de estrés universal. Función parcialmente desconocida. Se sabe que se une a ATP y que esta afinidad le permite reducir la proliferación de Mt in vitro y que permite a Mt mantenerse prolongadamente en estado latente in vivo. Se sugiere que esta proteína es una molécula de señalización mediada por ATP.
Esencialidad:	Esencial por mutagénesis transposomal Himar1 en H37Rv (in-vitro no esencial)
Metaboloma:	No hay data en estudio de FBA citado
Estructura:	PDB 3CIS. Cristalografía de rayos X con resolución de 2.9A y 3.2A (Max Planck Unit for Structural Molecular Biology, Hamburg Germany -publicación pendiente)
Cant. aa en pdb:	309
Grupo Hemo:	no posee
Ligandos:	Liga ATP e ion magnesio
Cu en cofactor:	no posee
Pfam data:	Usp (PF00582)
Cys/Tyr sitio activo:	no hay entrada en CSA

Proteína:	Bacterioferritin B (BfrB)
Gen Locus:	Rv3841
Función:	Almacenamiento de hierro y proceso de detoxificación.
Esencialidad:	No esencial por mutagénesis transposomal Himar1 en H37Rv
Metaboloma:	No hay data en estudio de FBA citado
Estructura:	PDB 3QD8/3UNO/3OJ5. Conformación homodímero de 24 monómeros y puede albergar hierro en su núcleo. Centro catalítico ferroxidativo. Cristalografía de rayos X con resolución de 3A (Univ. de Dehli South Campus, India)

Cant. aa en pdb:	189
Grupo Hemo:	no posee
Ligandos:	Liga iones de hierro
Cu en cofactor:	no posee
Pfam data:	Ferritin (PF00210)
Cys/Tyr sitio activo:	no hay entrada en CSA

Tabla 7. Información estructural, funcional y de esencialidad de las proteínas del TuberQuery Job con score mayor a 1

Si evaluamos la totalidad de las proteínas obtenidas a través de TuberQuery y los filtros aplicados, obtenemos que 9/23 genes obtenidos son identificados como esenciales.

La siguiente tabla muestra el detalle de todas las variables que se consideraron para la selección de blancos moleculares contra el Mt en estado de latencia. En verde se destacan las propiedades de las proteínas favorables para que la proteína en cuestión sea un candidato a blanco molecular de interés farmacológico. Las primeras seis entradas corresponden a los genes detallados en la tabla anterior.

En la columna PDB id se muestra solamente el PDB id con el mayor índice de drogabilidad de todas las estructuras PDB asociadas al ORF mediante el método descrito en este trabajo. El índice de drogabilidad puede ser utilizado para generar las siguientes categorías de proteínas: [ $<0.2$ ] no drogable, [ $0.2-0.5$ ] difícilmente drogable, [ $0.5-0.7$ ] drogable, [ $>0.7$ ] altamente drogable. El \* denota que la asociación Uniprot-PDB no existe en la base de datos Uniprot.

Locus Gen	PDB id (índice de drogabilidad)	Nombre Proteína	Nombre Recomendado	Esencial TraSH	Cant. de aa	Regulón dosR	Metabolismo FBA data	Metabolismo/Acción Anotado	Acción Catalítica	Cofactor (o Ligando)
Rv3133c	1ZLK (0.883)	devR dosR	Transcriptional regulatory protein DevR/DosR	no	217	si	no hay data	Factor de transcripción del dormancy regulon		
Rv2007c	2V2K (0.646) (*)	fdxA	Ferredoxin	no	114	si	no hay data	Transferencia electrones en reacciones metabólicas		4Fe-4S, 3Fe-4S
Rv2626c	1Y5H (0.642)	hrp1	Hypoxic response protein 1	no	143	si	no hay data	Posible rol en estado de latencia		Zn
Rv0569	2A7Y (0.939) (*)		Uncharacterized protein MT0595	no	88	si	no hay data	Posible rol en estado de latencia		
Rv2623	3CIS (0.856)		Universal stress protein MT2698	si	297	si	no hay data	Retarda el crecimiento en cultivo		ATP
Rv3841	3OJ5 (0.731)	bfrB	Ferritin BfrB	no	181	posible	no hay data	Oxida el hierro y mantiene homeostasis de hierro	$4 \text{Fe}^{2+} + 4 \text{H}^{+} + \text{O}_2 = 4 \text{Fe}^{3+} + 2 \text{H}_2\text{O}$	Fe
Rv0137c	1NWA (0.873)	msrA	Peptide methionine sulfoxide reductase MsrA	no	182	no	Other Amino Acid Metabolism	Repara proteínas inactivadas por oxidación	L-methionine + thioredoxin disulfide + $\text{H}_2\text{O} = \text{L-methionine (S)-S-oxide} + \text{thioredoxin}$	
Rv2539c	4BQS (0.737)	aroK	Shikimate kinase	si	176	no	Phenylalanine Tyrosine Tryptophan Metabolism	Foforila shikimate	$\text{ATP} + \text{shikimate} = \text{ADP} + \text{shikimate 3-phosphate}$	Mg
Rv2245	4C72 (0.777)	kasA	3-oxoacyl-[acyl-carrier-protein] synthase 1	si	416	no	Fatty Acid Metabolism	Cataliza la condensación de ácidos grasos	$\text{Acyl-[acyl-carrier-protein]} + \text{malonyl-[acyl-carrier-protein]} = 3\text{-oxoacyl-[acyl-carrier-protein]} + \text{CO}_2 + \text{[acyl-carrier-protein]}$	
Rv2540c	4BAJ (0.799)	aroC	Chorismate synthase	si	401	no	Phenylalanine Tyrosine Tryptophan Metabolism	Biosíntesis de chorismate	$5\text{-O-(1-carboxyvinyl)-3-phosphoshikimate} = \text{chorismate} + \text{phosphate}$	flavina reducida



Rv2246	2GP6 (0.709)	kasB	3-oxoacyl-[acyl-carrier-protein] synthase 2	si	417	no	Fatty Acid Metabolism	Cataliza la condensación de ácidos grasos	Acyl-[acyl-carrier-protein] + malonyl-[acyl-carrier-protein] = 3-oxoacyl-[acyl-carrier-protein] + CO <sub>2</sub> + [acyl-carrier-protein]	
Rv2737c	2ZRB (0.933)	recA	Recombinase A	si	790	no	no hay data	Une a DNA, hidroliza ATP y activa lexA		
Rv3290c	2JJH (0.906)	lat	Probable L-lysine-epsilon aminotransferase	no	449	no	Glutamate Metabolism		L-lysine + 2-oxoglutarate = (S)-2-amino-6-oxohexanoate + L-glutamate	pyridoxal fosfato
Rv2243	2QC3 (0.663)	fabD	Malonyl CoA-acyl carrier protein transacylase	no	302	no	Fatty Acid Metabolism	Metabolismo de lípidos	Malonyl-CoA + [acyl-carrier-protein] = CoA + malonyl-[acyl-carrier-protein]	
Rv2780	2VHV (0.879)	ald	Alanine dehydrogenase	no	371	no	Other Amino Acid Metabolism	Posible rol en pared celular en formación de pared celular (L-alanina)	L-alanine + H <sub>2</sub> O + NAD <sup>+</sup> = pyruvate + NH <sub>3</sub> + NADH	Mg
Rv0805	3IB8 (0.898)	cpdA	3',5'-cyclic adenosine monophosphate phosphodiesterase	no	318	no	Purine Metabolism	Regula concentración interna de cAMP	Adenosine 3',5'-cyclic phosphate + H <sub>2</sub> O = adenosine 5'-phosphate	Fe <sup>3+</sup> , Mn <sup>2+</sup>
Rv3709c	3S1T (0.486)	ask	Aspartokinase	si	421	no	Glycine, Serine, and Threonine Metabolism	Biosíntesis de aminoácidos	ATP + L-aspartate = ADP + 4-phospho-L-aspartate	
Rv3608c	1EYE (0.947)	folP1	Dihydropteroate synthase 1	si	280	no	Folate Metabolism		(2-amino-4-hydroxy-7,8-dihydropteridin-6-yl)methyl diphosphate + 4-aminobenzoate = diphosphate + dihydropteroate	Mg (requerido)
Rv0859	3SVK (0.850)	fadA	Putative acyltransferase	no	403	no	Fatty Acid Metabolism			

Rv0046c	1GR0 (0.719)	ino1	Inositol-3-phosphate synthase	no	367	no	Peptidoglycan Metabolism		D-glucose 6-phosphate = 1D-myo-inositol 3-phosphate	
Rv1177	2V2K (0.646) (*)	fdxC	Ferredoxin	no	108	no	no hay data			4Fe-4S
Rv2069	2O7G (0.881)	sigC	ECF RNA polymerase sigma factor	no	185	no	no hay data	Factor de iniciación durante la transcripción		AD.N
Rv3581c	2UZH (no data) (*)	ispF	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	si	159	no	Polyprenyl Metabolism		2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate + CMP	Metales divalentes

Tabla 8. Resumen de información estructural, funcional, metabólica, y de esencialidad de las proteínas del TuberQuery Job con score mayor o igual a 1

A continuación se muestra un resumen de resultados de genes obtenidos luego de cada uno de los pasos desarrollados en esta tesina.

Datos	Fuente
4018 ORFs (40Mbp) 25-50% función desconocida	Genoma de Mt, TubercuList
244 genes	Sobre-expresados en condiciones de stress >200% según citas seleccionadas
23 proteínas únicas (52 tripletes únicos UNIPROT-PFAM-PDB)	Selección anterior perteneciente a familia Pfam y estructura relacionada en PDB según TuberQuery
9 proteínas	Selección anterior con esencialidad según citas de TraSH analizadas
7 proteínas	Selección anterior con índice de drogabilidad altamente drogable según criterios en TuberQuery

Tabla 9. Resumen de candidatos a blanco molecular tras cada filtro de selección

Adicionalmente, de las 7 proteínas obtenidas tras la aplicación de todos los filtros, 5 de ellas están representados en los análisis de FBA y actúan en vías metabólicas. Las dos proteínas que no están representadas en los estudios de FBA son recA y MT2698. RecA es una proteína de reparación de ADN común a muchos organismos. MT2698 es una proteína desconocida en su función. A continuación se muestra la captura de pantalla de MT2698 en TuberQuery. En esta captura el bolsillo más drogable de la proteína en estado monomérico tiene un índice de drogabilidad de 0.526 (cavidad marcada con esferas negras en el gráfico). En su estructura cuaternaria la macromolécula se presenta como un tetrámero con un bolsillo con índice de drogabilidad de 0.856.

006189 [10-148]:UNIVERSAL STRESS PROTEIN RV2623/MT2698| PF00582.21 [3CIS]

Initials	3cis_a Summary Structure					
3CIS	Structure has pKa >= 1 from reference	true	Max Pocket Druggability	0.856	Structure has drug	false
3cis_a	Pocket Matches PFAM relevant residue	true	Structure is in CSA	false	Structure has Metals	true

Metadata

Data Visualization Panel

Select	Druggability	Drug/CSA
<input checked="" type="checkbox"/>	0.526	No / No
<input type="checkbox"/>	0.267	No / No
<input type="checkbox"/>	0.242	No / No

Cavities computed by fpocket 2.0. Homology model generated with Modeller 2. Pka computed by propka 3.1.

[Link to RSCB PDB](#)
[Pockets: 9](#)
[Metals: 0](#)
[propKa: 0](#)
[CSA: 0](#)
[Ligands: 0](#)

Gráfico 4. Captura de pantalla de TuberQuery para estructura PDB 3CIS asociada al gen locus MT2698

#### 4 DISCUSIÓN DE RESULTADOS

Se realizó la primera fase de una selección racional de blancos moleculares de potencial interés para el desarrollo de nuevos fármacos en la lucha contra la tuberculosis, con el particular enfoque de atacar al bacilo de la TB en su estado de latencia y así disminuir el reservorio mundial de Mt latente.

Se priorizó durante esta tesina la conceptualización del flujo de trabajo para dicho objetivo. Por lo tanto no consideramos la lista de candidatos obtenida como una lista robusta y exhaustiva. Esto requeriría entre otras cosas:

- La ampliación de la cantidad de trabajos científicos de sobre-expresión analizados con diversos enfoques para tratar de simular las condiciones de latencia reales en el huésped: hipoxia in-vitro, RNOS in-vitro, in-vivo, etc.
- La asignación de un peso específico a cada microarreglo analizado en base a su robustez estadística y su relevancia de acuerdo a los objetivos del trabajo.
- La asignación de un valor de sensibilidad de detección de sobre-expresión que considere no solamente el *fold-change* sino también los valores absolutos de expresión de cada proteína y del conjunto del microarreglo.
- Una iteración de búsquedas para optimizar los valores de los parámetros especificados a TuberQuery y encontrar el mejor equilibrio entre sensibilidad y especificidad.

- Una asignación de valores que considere la función metabólica de los *hits* en base a la bioquímica celular del patógeno y su micro-entorno en el huésped.
- Un análisis estadístico para mostrar que la cantidad de blancos moleculares encontrados por este método es significativamente diferente al número que se encontraría por azar dada la cantidad de datos analizados.
- Extender el filtro de esencialidad vía ensayos de hibridación por transposones y análisis in-silico de balance de flujo de vías metabólicas a la totalidad de genes sobre-expresados obtenidos sin aplicar filtros relacionados a estructura para obtener así un listado menos restrictivo de genes con altas posibilidades de ser bactericidas o bacteriostático para la Mt latente. Estos serían genes esenciales sobre-expresados durante la TB latente con o sin información estructural disponible.
- Incluir en este estudio las proteínas sub-expresadas que pueden ser fuente de valiosa información respecto a la adaptación y supervivencia del patógeno en el estado de latencia. Podría ocurrir que alguna de estas proteínas adquiriera cierta sensibilidad durante el estado de latencia y se convirtiera en un blanco molecular drogable contra la TB latente.

La focalización de microarreglos en condiciones de hipoxia y la aplicación del filtro estructural con TuberQuery pudo haber sido demasiado restrictivo y que hayamos perdido información valiosa en la búsqueda de blancos moleculares contra la TB como lo demuestra la falta de detección de las proteínas que a

priori esperábamos como controles positivos, trHbN - Rv1542c (M. A. Martí et al. 2008) y Cyp121 - Rv2276 (Dumas et al. 2013). Ninguna de estas dos proteínas fue identificada como sobre-expresada en los microarregles analizados en este trabajo y con los criterios utilizados. No obstante hay otros microarreglos de Mt en estado de latencia que detectan a Cyp121 como sub-expresado y a trHbN como sobre-expresado. Adicionalmente, si bien trHbN no fue predicha como blanco molecular en este trabajo, esta molécula fue cristalizada con Isoniazida, lo cual demuestra su *bindability*. Cabe destacar que no existen drogas comerciales específicas para atacar a la Mt en fase de latencia con lo cual no contamos con controles positivos de identificación inmediata. Pero se debería realizar una búsqueda bibliográfica más profunda para la determinación y ampliación de los potenciales controles positivos.

En base a los resultados obtenidos podemos destacar que el análisis de sobre-expresión genómica en condiciones de estrés arrojó como resultado una gran cantidad de genes pertenecientes al regulón DosR. Cuando se utilizó la base de datos TuberQuery la cantidad de genes del regulón DosR encontrados se redujo significativamente. Es posible que esto se deba a la falta de identificación de dominios Pfam en las proteínas putativas del regulón DosR sobre-expresadas y/o a la falta de estructuras cristalizadas para dichas familias de proteínas. Esto es un tema a investigar en mayor profundidad.

Creemos que la búsqueda y desarrollo de inhibidores de proteínas de Mt relevantes a la fase de latencia, seleccionadas en base a un análisis contextual basado en información de secuencia-estructura y esencialidad, es una buena

estrategia para atacar el reservorio de Mt latente que es uno de los causales del desarrollo de cepas MDR/XDR y de la alta mortandad causada por esta patología.

Proponemos que la evaluación estructural de proteínas homologas o proteínas que comparten dominios conservados es una alternativa efectiva para estudiar y ponderar la potencialidad farmacológica de aquellas proteínas confirmadas o hipotéticas de estructura desconocida y de interés en la patología latente de TB.

Por último proponemos que el método de búsqueda racional bioinformático de blancos moleculares utilizado para esta tesina y continuado en el PICT 2010-2805 (M. Martí and Turjanski 2010) permite al investigador aumentar las posibilidades de éxito en búsqueda de nuevos fármacos optimizando la utilización de recursos que este objetivo requiere.

Por consiguiente consideramos que esta es una herramienta útil para reducir el tiempo y capital necesarios para el descubrimiento de nuevos blancos moleculares en las etapas iniciales de investigación de nuevos fármacos en general, y en particular para aquellas patologías que por diferentes motivos socio-económicos, geopolíticos o de financiación no logran ser curadas o controladas al máximo de las capacidades científicas corrientes.

## 5 TRABAJO FUTURO

Adicionalmente a los puntos expuestos en la sección anterior a fortalecer para obtener una lista exhaustiva y robusta de proteínas candidatas a ser blancos moleculares, a continuación enumeramos algunas de las tareas que se deberían afrontar como trabajo a futuro.

Una vez identificado el conjunto de genes esenciales y sobre-expresados en condiciones patológicas relevantes se debe continuar con una profundización del estudio estructural para determinar la drogabilidad de las proteínas candidatas a blancos moleculares. En esta etapa también se puede analizar la conformidad a diferentes estándares predictivos de drogabilidad como la regla de cinco de Lipinski (Petit et al. 2012; Pollastri 2010).

Una vez obtenidas las proteínas hipotéticamente esenciales, sobre-expresadas en estado de latencia, y drogables, se debe proceder a la validación del blanco molecular. Se puede confirmar la sobre-expresión del blanco molecular mediante otros métodos experimentales como PCR cuantitativo de tiempo real (qRT-PCR) en las condiciones de experimentación pertinentes. También se puede confirmar la esencialidad de la proteína realizando un experimento de knock-out o knock-out selectivo.

El paso siguiente sería la realización de ensayos de modelado dinámico de proteínas y escaneo virtual de librerías de drogas sobre los blancos molecular seleccionados para identificar los compuestos líderes que mejor inhiban al



target. Luego proceder a la optimización racional *in-silico* de estos compuestos líderes para lograr una mayor afinidad y especificidad de unión/inhibición con los blancos moleculares en cuestión.

Una vez en posesión de las drogas selectivas, se puede poner a punto un ensayo de viabilidad celular. El grupo de investigación del Dr. Martí, director de esta tesis, cuenta con un ensayo de multiplacas de *Mycobacterium smegmatis* (Ms) en un laboratorio de bioseguridad 2. En estas placas se cultivan células de Ms y se hacen transfecciones con un micobacteriófago con marcación via GFP que permite expresar selectivamente la proteína blanco. La droga puede administrarse en paralelo con el micobacteriófago o posteriormente dependiendo del mecanismo de acción de la misma, y ver si la Ms sobrevive o no, y si expresa el marcador GFP o no. Con este ensayo se pueden realizar las curvas de respuesta a diferentes dosis de la droga, para diferentes candidatos a droga y/o blancos moleculares. También se deben realizar ensayos *in-vitro* de farmacocinética y farmacodinamia como es la determinación de las constantes de unión/inhibición de las drogas y el blanco molecular seleccionado.

Una vez finalizados los ensayos *in-vitro* y confirmada de la acción de la droga sobre Ms habría que continuar con experimentos *in-vivo* con cepas de Mt en condiciones de latencia en laboratorios con protocolos y seguridad para dichos experimentos. Durante la etapa de estudios en animales se continúan los ensayos de farmacocinética y farmacodinamia para determinar los criterios ADME (adsorción, distribución, metabolismo, secreción), la biodisponibilidad oral, la

toxicidad en el huésped, y la efectividad in vivo de las drogas seleccionadas contra el blanco molecular identificado.

## **6 AGRADECIMIENTOS**

Adrian Turjanski \*

Lucas A. Defelipe \*

Leandro Radusky \*

\* Grupo de Bioinformática Estructural, FCEyN, UBA.

Fundación Bunge y Born para investigación sobre enfermedades infecciosas.

Proyecto de Investigación Científica y Tecnológica (PICT) BICENTENARIO 2010-2805 de la Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT) del Ministerio de Ciencia, Tecnología e Innovación Productiva (MINCyT) (M. Martí and Turjanski 2010).

Cámara de Empresas de software y Servicios Informáticos de Argentina (CESSI) y la Fundación Sadosky por el Premio Sadosky 2013 al desarrollo de la herramienta TuberQuery en la categoría Trabajo de Investigación Colaborativo Industria – Academia junto con la empresa BioFlux Bioinformática.

## 7 REFERENCIAS

- Agüero, Fernán, Bissan Al-Lazikani, Martin Aslett, Matthew Berriman, Frederick S Buckner, Robert K Campbell, Santiago Carmona, et al. 2008. “Genomic-Scale Prioritization of Drug Targets: The TDR Targets Database.” *Nature Reviews. Drug Discovery* 7 (11): 900–907. doi:10.1038/nrd2684.
- Alonso, Hernán, Andrey A Bliznyuk, and Jill E Gready. 2006. “Combining Docking and Molecular Dynamic Simulations in Drug Design.” *Medicinal Research Reviews* 26 (5): 531–68. doi:10.1002/med.20067.
- Anishetty, Sharmila, Mrudula Pulimi, and Gautam Pennathur. 2005. “Potential Drug Targets in Mycobacterium Tuberculosis through Metabolic Pathway Analysis.” *Computational Biology and Chemistry* 29 (5): 368–78. doi:10.1016/j.compbiolchem.2005.07.001.
- Berman, H M, J Westbrook, Z Feng, G Gilliland, T N Bhat, H Weissig, I N Shindyalov, and P E Bourne. 2000. “The Protein Data Bank.” *Nucleic Acids Research* 28 (1): 235–42.
- Beste, Dany J V, Tracy Hooper, Graham Stewart, Bhushan Bonde, Claudio Avignone-Rossa, Michael E Bushell, Paul Wheeler, Steffen Klamt, Andrzej M Kierzek, and Johnjoe McFadden. 2007. “GSMN-TB: A Web-Based Genome-Scale Network Model of Mycobacterium Tuberculosis Metabolism.” *Genome Biology* 8 (5): R89. doi:10.1186/gb-2007-8-5-r89.
- “Biocyc Database Collection.” 2013. Accessed December 2. <http://biocyc.org/>.
- Blundell, Tom L, Bancinyane L Sibanda, Rinaldo Wander Montalvão, Suzanne Brewerton, Vijayalakshmi Chelliah, Catherine L Worth, Nicholas J Harmer, Owen Davies, and David Burke. 2006. “Structural Biology and Bioinformatics in Drug Design: Opportunities and Challenges for Target Identification and

- Lead Discovery.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 361 (1467): 413–23. doi:10.1098/rstb.2005.1800.
- Bonora, Stefano, and Giovanni Di Perri. 2008. “Interactions between Antiretroviral Agents and Those Used to Treat Tuberculosis.” *Current Opinion in HIV and AIDS* 3 (3): 306–12. doi:10.1097/COH.0b013e3282fbaad0.
- “Centers for Disease Control and Prevention (CDC).” 2013. Accessed June 3. <http://www.cdc.gov/tb/>.
- Chen, Yi-Ping Phoebe, and Feng Chen. 2008a. “Identifying Targets for Drug Discovery Using Bioinformatics.” *Expert Opinion on Therapeutic Targets* 12 (4): 383–89. doi:10.1517/14728222.12.4.383.
- . 2008b. “Using Bioinformatics Techniques for Gene Identification in Drug Discovery and Development.” *Current Drug Metabolism* 9 (6): 567–73.
- “Clinicaltrials.gov Search Results.” 2012. *Search of: Tuberculosis - List Results - ClinicalTrials.gov*. Accessed September 9. <http://clinicaltrials.gov/ct2/results?term=tuberculosis>.
- Cole, S T. 1999. “Learning from the Genome Sequence of Mycobacterium Tuberculosis H37Rv.” *FEBS Letters* 452 (1-2): 7–10.
- Cole, S T, R Brosch, J Parkhill, T Garnier, C Churcher, D Harris, S V Gordon, et al. 1998. “Deciphering the Biology of Mycobacterium Tuberculosis from the Complete Genome Sequence.” *Nature* 393 (6685): 537–44. doi:10.1038/31159.
- Dark, Michael J. 2013. “Whole-Genome Sequencing in Bacteriology: State of the Art.” *Infection and Drug Resistance* 6: 115–23. doi:10.2147/IDR.S35710.
- Draghici, Sorin, Purvesh Khatri, Adi Laurentiu Tarca, Kashyap Amin, Arina Done, Calin Voichita, Constantin Georgescu, and Roberto Romero. 2007. “A Systems

- Biology Approach for Pathway Level Analysis.” *Genome Research* 17 (10): 1537–45. doi:10.1101/gr.6202607.
- Dumas, Victoria G, Lucas A Defelipe, Ariel A Petruk, Adrian G Turjanski, and Marcelo A Marti. 2013. “QM/MM Study of the C-C Coupling Reaction Mechanism of CYP121, an Essential Cytochrome p450 of Mycobacterium Tuberculosis.” *Proteins*, November. doi:10.1002/prot.24474.
- Dye, C, S Scheele, P Dolin, V Pathania, and M C Raviglione. 1999. “Consensus Statement. Global Burden of Tuberculosis: Estimated Incidence, Prevalence, and Mortality by Country. WHO Global Surveillance and Monitoring Project.” *JAMA: The Journal of the American Medical Association* 282 (7): 677–86.
- Fine, P E. 1995. “Variation in Protection by BCG: Implications of and for Heterologous Immunity.” *Lancet* 346 (8986): 1339–45.
- Gao, Qian, Katharine E Kripke, Alok J Saldanha, Weihong Yan, Susan Holmes, and Peter M Small. 2005. “Gene Expression Diversity among Mycobacterium Tuberculosis Clinical Isolates.” *Microbiology (Reading, England)* 151 (Pt 1): 5–14. doi:10.1099/mic.0.27539-0.
- “Global Tuberculosis Control -Epidemiology, Strategy, Financing.” 2009. WHO/HTM/TB/2009.411. [http://www.who.int/tb/publications/global\\_report/2009/en/index.html](http://www.who.int/tb/publications/global_report/2009/en/index.html).
- Griffin, Jennifer E, Jeffrey D Gawronski, Michael A Dejesus, Thomas R Ioerger, Brian J Akerley, and Christopher M Sassetti. 2011. “High-Resolution Phenotypic Profiling Defines Genes Essential for Mycobacterial Growth and Cholesterol Catabolism.” *PLoS Pathogens* 7 (9): e1002251. doi:10.1371/journal.ppat.1002251.

- Hampshire, Tobias, Shamit Soneji, Joanna Bacon, Brian W. James, Jason Hinds, Ken Laing, Richard A Stabler, Philip D. Marsh, and Philip D Butcher. 2004. "Stationary Phase Gene Expression of Mycobacterium Tuberculosis Following a Progressive Nutrient Depletion: A Model for Persistent Organisms?" *Tuberculosis* 84 (3-4): 228–38. doi:10.1016/j.tube.2003.12.010.
- Hasan, Samiul, Sabine Daugelat, P. S. Srinivasa Rao, and Mark Schreiber. 2006. "Prioritizing Genomic Drug Targets in Pathogens: Application to Mycobacterium Tuberculosis." *PLoS Computational Biology* 2 (6): e61. doi:10.1371/journal.pcbi.0020061.
- Irwin, John J, and Brian K Shoichet. 2005. "ZINC--a Free Database of Commercially Available Compounds for Virtual Screening." *Journal of Chemical Information and Modeling* 45 (1): 177–82. doi:10.1021/ci049714+.
- Jamshidi, Neema, and Bernhard Ø Palsson. 2007. "Investigating the Metabolic Capabilities of Mycobacterium Tuberculosis H37Rv Using the in Silico Strain iNJ661 and Proposing Alternative Drug Targets." *BMC Systems Biology* 1: 26. doi:10.1186/1752-0509-1-26.
- Janin, Yves L. 2007. "Antituberculosis Drugs: Ten Years of Research." *Bioorganic & Medicinal Chemistry* 15 (7): 2479–2513. doi:10.1016/j.bmc.2007.01.030.
- Jorgensen, William L. 2004. "The Many Roles of Computation in Drug Discovery." *Science (New York, N.Y.)* 303 (5665): 1813–18. doi:10.1126/science.1096361.
- . 2009. "Efficient Drug Lead Discovery and Optimization." *Accounts of Chemical Research* 42 (6): 724–33. doi:10.1021/ar800236t.
- "KEGG Pathway Database." 2013. Accessed April 3. <http://www.genome.jp/kegg/pathway.html>.

- Malone, John H, and Brian Oliver. 2011. "Microarrays, Deep Sequencing and the True Measure of the Transcriptome." *BMC Biology* 9: 34. doi:10.1186/1741-7007-9-34.
- Mandal, Soma, Mee'nal Moudgil, and Sanat K Mandal. 2009. "Rational Drug Design." *European Journal of Pharmacology* 625 (1-3): 90–100. doi:10.1016/j.ejphar.2009.06.065.
- Martí, Marcelo A, Axel Bidon-Chanal, Alejandro Crespo, Syun-Ru Yeh, Victor Guallar, F Javier Luque, and Darío A Estrin. 2008. "Mechanism of Product Release in NO Detoxification from Mycobacterium Tuberculosis Truncated Hemoglobin N." *Journal of the American Chemical Society* 130 (5): 1688–93. doi:10.1021/ja076853+.
- Martí, Marcelo, and Adrian Turjanski. 2010. "PICT 2010-2805". Grupo de Bioinformática Estructural, Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires.
- Muttucumaru, D G Niranjala, Gretta Roberts, Jason Hinds, Richard A Stabler, and Tanya Parish. 2004. "Gene Expression Profile of Mycobacterium Tuberculosis in a Non-Replicating State." *Tuberculosis (Edinburgh, Scotland)* 84 (3-4): 239–46. doi:10.1016/j.tube.2003.12.006.
- Nguyen, Liem, and Jean Pieters. 2009. "Mycobacterial Subversion of Chemotherapeutic Reagents and Host Defense Tactics: Challenges in Tuberculosis Drug Development." *Annual Review of Pharmacology and Toxicology* 49: 427–53. doi:10.1146/annurev-pharmtox-061008-103123.
- Nwaka, Solomon, and Robert G Ridley. 2003. "Virtual Drug Discovery and Development for Neglected Diseases through Public-Private Partnerships." *Nature Reviews. Drug Discovery* 2 (11): 919–28. doi:10.1038/nrd1230.

- Ohno, Hideaki, Guofeng Zhu, Vellore P. Mohan, Darien Chu, Shigeru Kohno, William R. Jacobs, and John Chan. 2003. "The Effects of Reactive Nitrogen Intermediates on Gene Expression in *Mycobacterium Tuberculosis*." *Cellular Microbiology* 5 (9): 637–48. doi:10.1046/j.1462-5822.2003.00307.x.
- Ortí, Leticia, Rodrigo J Carbajo, Ursula Pieper, Narayanan Eswar, Stephen M Maurer, Arti K Rai, Ginger Taylor, et al. 2009. "A Kernel for the Tropical Disease Initiative." *Nature Biotechnology* 27 (4): 320–21. doi:10.1038/nbt0409-320.
- Palmer, C E, and M W Long. 1966. "Effects of Infection with Atypical Mycobacteria on BCG Vaccination and Tuberculosis." *The American Review of Respiratory Disease* 94 (4): 553–68.
- Park, Heui-Dong, Kristi M. Guinn, Maria I. Harrell, Reiling Liao, Martin I. Voskuil, Martin Tompa, Gary K. Schoolnik, and David R. Sherman. 2003. "Rv3133c/dosR Is a Transcription Factor That Mediates the Hypoxic Response of *Mycobacterium Tuberculosis*." *Molecular Microbiology* 48 (3): 833–43.
- Petit, Joachim, Nathalie Meurice, Christine Kaiser, and Gerald Maggiora. 2012. "Softening the Rule of Five--Where to Draw the Line?" *Bioorganic & Medicinal Chemistry* 20 (18): 5343–51. doi:10.1016/j.bmc.2011.11.064.
- Pollastri, Michael P. 2010. "Overview on the Rule of Five." *Current Protocols in Pharmacology / Editorial Board, S.J. Enna (editor-in-Chief) ... [et Al.]* Chapter 9 (June): Unit 9.12. doi:10.1002/0471141755.ph0912s49.
- "Pubmed Search Results 1." 2012. "*Mycobacterium Tuberculosis*" AND "Microarray" AND "Latent" - PubMed - NCBI. Accessed August 27.  
<http://www.ncbi.nlm.nih.gov/pubmed?term=%E2%80%9CMycobacterium%20tuberculosis%E2%80%9D%20AND%20%22microarray%22%20AND%20%22latent%22>.



- “Pubmed Search Results 2.” 2012. “*Tuberculosis*” AND “*Microarray*” AND “*Latent*” - *PubMed - NCBI*. Accessed August 27.
- <http://www.ncbi.nlm.nih.gov/pubmed?term=%22tuberculosis%22%20AND%20%22microarray%22%20AND%20%22latent%22>.
- Raman, Karthik, Preethi Rajagopalan, and Nagasuma Chandra. 2005. “Flux Balance Analysis of Mycolic Acid Pathway: Targets for Anti-Tubercular Drugs.” *PLoS Computational Biology* 1 (5): e46. doi:10.1371/journal.pcbi.0010046.
- “Reporte Sobre Incidencia Y Mortalidad Por Tuberculosis En Argentina.” 2008. Nacion, M. d. S. d. l.
- Rester, Ulrich. 2008. “From Virtuality to Reality - Virtual Screening in Lead Discovery and Lead Optimization: A Medicinal Chemistry Perspective.” *Current Opinion in Drug Discovery & Development* 11 (4): 559–68.
- Sasseti, Christopher M, Dana H Boyd, and Eric J Rubin. 2003. “Genes Required for Mycobacterial Growth Defined by High Density Mutagenesis.” *Molecular Microbiology* 48 (1): 77–84.
- Sauer, Robert. 2005. *Reducing Barriers to the Development of High Quality, Low Cost Medicines: A Proposal for Reforming the Drug Approval Process*. International Policy Press.
- Scapin, Giovanna. 2006. “Structural Biology and Drug Discovery.” *Current Pharmaceutical Design* 12 (17): 2087–97.
- Schnappinger, D., S. Ehrt, M. I. Voskuil, Y. Liu, J. A. Mangan, I. M. Monahan, G. Dolganov, et al. 2003. “Transcriptional Adaptation of Mycobacterium Tuberculosis within Macrophages: Insights into the Phagosomal Environment.” *Journal of Experimental Medicine* 198 (5): 693–704.
- doi:10.1084/jem.20030846.

- Seco, Jesus, F Javier Luque, and Xavier Barril. 2009. "Binding Site Detection and Druggability Index from First Principles." *Journal of Medicinal Chemistry* 52 (8): 2363–71. doi:10.1021/jm801385d.
- Sherman, David R., Martin Voskuil, Dirk Schnappinger, Reiling Liao, Maria I. Harrell, and Gary K. Schoolnik. 2001. "Regulation of the Mycobacterium Tuberculosis Hypoxic Response Gene Encoding A-Crystallin." *Proceedings of the National Academy of Sciences* 98 (13): 7534–39.
- Taylor, R D, P J Jewsbury, and J W Essex. 2002. "A Review of Protein-Small Molecule Docking Methods." *Journal of Computer-Aided Molecular Design* 16 (3): 151–66.
- "The Kyoto Encyclopedia of Genes and Genomes—KEGG." 2000. *Yeast* 1 (1): 48–55. doi:10.1002/(SICI)1097-0061(200004)17:1<48::AID-YEA2>3.0.CO;2-H.
- Tramontano, Anna. 2006. "The Role of Molecular Modelling in Biomedical Research." *FEBS Letters* 580 (12): 2928–34. doi:10.1016/j.febslet.2006.04.011.
- "Tuberquery." 2014. Accessed February 10. [tuberq.proteinq.com.ar](http://tuberq.proteinq.com.ar).
- "Unión Internacional Contra La Tuberculosis Y Enfermedades Respiratorias." <http://www.union-imdp.org/files/90th%20broch%20Sp%20LR.pdf>.
- Venkataswamy, Manjunatha M, Michael F Goldberg, Andres Baena, John Chan, William R Jacobs Jr, and Steven A Porcelli. 2012. "In Vitro Culture Medium Influences the Vaccine Efficacy of Mycobacterium Bovis BCG." *Vaccine* 30 (6): 1038–49. doi:10.1016/j.vaccine.2011.12.044.
- Voskuil, M I, K C Visconti, and G K Schoolnik. 2004. "Mycobacterium Tuberculosis Gene Expression during Adaptation to Stationary Phase and Low-Oxygen Dormancy." *Tuberculosis (Edinburgh, Scotland)* 84 (3-4): 218–27. doi:10.1016/j.tube.2004.02.003.

- Voskuil, M. I., D. Schnappinger, K. C. Visconti, M. I. Harrell, G. M. Dolganov, D. R. Sherman, and G. K. Schoolnik. 2003. "Inhibition of Respiration by Nitric Oxide Induces a Mycobacterium Tuberculosis Dormancy Program." *Journal of Experimental Medicine* 198 (5): 705–13. doi:10.1084/jem.20030205.
- Vosmeer, C Ruben, René Pool, Mariël F Van Stee, Lovorka Peric-Hassler, Nico P E Vermeulen, and Daan P Geerke. 2014. "Towards Automated Binding Affinity Prediction Using an Iterative Linear Interaction Energy Approach." *International Journal of Molecular Sciences* 15 (1): 798–816. doi:10.3390/ijms15010798.
- Walters, W P, Ajay, and M A Murcko. 1999. "Recognizing Molecules with Drug-like Properties." *Current Opinion in Chemical Biology* 3 (4): 384–87.
- Wayne, L G, and G A Diaz. 1967. "Autolysis and Secondary Growth of Mycobacterium Tuberculosis in Submerged Culture." *Journal of Bacteriology* 93 (4): 1374–81.

## **FIRMAS DE LOS RESPONSABLES Y PARTICIPANTES DEL PROYECTO**

.....  
Ivan M. Kruk, Maestrando

.....  
Marcelo A. Martí, Director