Biblioteca Digital
F C E N - U B A

BIBLIOTECA CENTRAL LUIS F LELOIR
BIBLIOTECA CENTRAL LELOIR
FACULTAD DE CIENCIAS EXACTAS Y NATURALES UBA

Tesis Doctoral

Nuevas estrategias computacionales para la elección de blancos y desarrollo de nuevos inhibidores en Mycobacterium tuberculosis

Defelipe, Lucas Alfredo

2016-03-30

Este documento forma parte de la colección de tesis doctorales y de maestría de la Biblioteca Central Dr. Luis Federico Leloir, disponible en digital.bl.fcen.uba.ar. Su utilización debe ser acompañada por la cita bibliográfica con reconocimiento de la fuente.

This document is part of the doctoral theses collection of the Central Library Dr. Luis Federico Leloir, available in digital.bl.fcen.uba.ar. It should be used accompanied by the corresponding citation acknowledging the source.

Cita tipo APA:
Defelipe, Lucas Alfredo. (2016-03-30). Nuevas estrategias computacionales para la elección de blancos y desarrollo de nuevos inhibidores en Mycobacterium tuberculosis. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.

Cita tipo Chicago:
Defelipe, Lucas Alfredo. "Nuevas estrategias computacionales para la elección de blancos y desarrollo de nuevos inhibidores en Mycobacterium tuberculosis". Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2016-03-30.

EXACTAS UBA
Facultad de Ciencias Exactas y Naturales

UBA
Universidad de Buenos Aires

Dirección: Biblioteca Central Dr. Luis F. Leloir, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires. Intendente Güiraldes 2160 - C1428EGA - Tel. (++54 +11) 4789-9293

Contacto: digital@bl.fcen.uba.ar

UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Química Biológica

Nuevas estrategias computacionales para la elección de blancos y desarrollo de nuevos inhibidores en *Mycobacterium tuberculosis*

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

Lucas Alfredo Defelipe

Directores de tesis: Adrián Gustavo Turjanski

Marcelo Adrián Martí

Consejero de estudios: Valeria Levi

Lugar de trabajo: Departamento de Química Biológica, FCEN-UBA e INQUIMAE-UBA/CONICET

Buenos Aires, Marzo de 2016

Fecha de defensa 30/03/2016

0.1 Resumen

La tuberculosis sigue siendo un problema de salud a nivel mundial. Un millón y medio de personas al año mueren por esta enfermedad siendo la primera causa de muerte entre los infectados con el virus de la inmunodeficiencia humana (HIV). El microorganismo causante de la enfermedad, *Mycobacterium tuberculosis (Mtb.)* , es una bacteria de crecimiento lento que vive dentro de los macrófagos del hospedador, en donde puede residir por años sin producir ningún síntoma de la infección en un estado que se conoce como de **latencia**. Dentro del macrófago la mycobacteria se encuentra sometida a una serie de condiciones de estrés como son la hipoxia, la falta de nutrientes y la presencia de especies reactivas de oxígeno y nitrógeno (ERON). *Mtb* dispone de varios mecanismos de protección en dicho ambiente de los que se vale para sobrevivir durante años.

El presente trabajo de tesis tiene como principal objetivo proponer nuevos procedimientos para el descubrimiento de blancos relevantes para la fase latente de *Mtb.* Se encuentra dividido en tres partes: La primera parte trata sobre el armado de una base de datos de proteínas particular de *Mycobacterium tuberculosis*, llamada TuberQ. Dicha base de datos cuenta con información genómica contextual (expresión en diversas condiciones que simulan el estrés durante la vida dentro del macrófago), información estructural retirada del RSCB PDB o genera a partir de modelado comparativo, cálculos de drogabilidad estructural, determinaciones de sensibilidad a estrés por la presencia de residuos cisteína/tirosina o la presencia de centros metálicos oxidables (Fe, Cu, Zn) y una reconstrucción de los principales metabolismos de *Mtb*. Combinamos esta información mediante el uso de una función de puntuación para clasificar los blancos moleculares respecto a su relevancia como blancos terapéuticos contra *Mtb.* en condiciones de estrés. Se clasificaron de esta forma miles de proteínas, en particular proteínas novedosas pertenecientes a la vía de síntesis de micotiol) o blancos ya descriptos (como la vía de síntesis de ácido micólico).

En la segunda parte nos encargamos de realizar un estudio mediante técnicas bioinformáticas y de química computacional de la familia de proteínas perteneciente a las *Cyclopropane Mycolic Acid Synthases (CMAS)*, dichas enzimas forman parte de la vía de síntesis de ácido micólico y se encuentran validadas como potenciales blancos terapeúticos. Estas enzimas son metiltransferasas dependientes de S-adenosil-L-metionina (SAM) con una particularidad, cada una posee

una selectividad marcada y un producto específico pudiendo realizar reacciones tan diversas como la ciclopropilación o generar grupos metil-alcoholes a partir de olefinas. Encontramos los determinantes moleculares de la selectividad de los productos en las distintas CMAS, proponiendo algunos cambios puntuales para poder validar estos resultados experimentalmente.

En la última parte de este trabajo de tesis nos abocamos a diseñar un protocolo de búsqueda virtual de compuestos específicos para la subfamilia de enzimas CMAS que contienen un ión bicarbonato en su sitio activo (cmaA1-2, pcaA, mmaA2 y umaA) utilizando a umaA como referencia. En esta sección proponemos distintos grupos miméticos al bicarbonato (Urea, Glicina, Carbamato, Ácido metoxiacético y Hidroxifuranona) que pueden actuar como farmacóforos buscando en la base de datos ZINC. Esta base de datos de más de 30.000 compuestos fue utilizada para realizar la búsqueda virtual con una posterior estimación de las energías de unión mediante dinámica molecular y MM-PBSA. Concluimos que los mejores compuestos están basados en Urea y Glicina debido, en parte, a la interacción que pueden realizar con sus grupos amino con un glutámico presente en el sitio activo de este subconjunto de CMAS.

En resumen, hemos generado un *pipeline* bioinformático que permite la elección de blancos con los criterios que el usuario prefiera, probado mecanismos de reacción de las CMAS y utilizado esta información para proponer nuevas moléculas con actividad bactericida.

0.2 Abstract

Tuberculosis remains a worldwide issue, 1.5 million people die from TB each year being the first cause of death among HIV infected people. The microorganism responsible for TB, *Mycobacterium tuberculosis*, is a slow growing bacteria which lives inside host's macrophages where it can stay for years without causing symptoms to the host, in a state known as **latency**. Inside the macrophage, mycobacteria is exposed to a series of stressful conditions such as hypoxia, starvation or the presence of reactive oxygen and nitrogen species (RNOS). *Mtb.* has various protection mechanisms in such environment which it uses to survive for years.

The present work main objective is to propose new procedures to discover latent phase relevant targets for *Mtb*. It is divided in three parts: The first one covers the design and setup of a protein database tailored for *Mycobacterium tuberculosis* called TuberQ. This database contains context-based genomic data (micro-array expression in infection-mimicking conditions as hypoxia, starvation and exposure to RNOS), structural information from RSCB PDB or derived by homology modeling, structural druggability computations, RNOS stress sensitivity by Cys/Tyr or metal center (Fe, Cu, Zn) containing proteins and a reconstruction of the principal metabolic pathways in *Mtb*. Using a scoring function we combine this information in order to classify the proteins regarding its relevance as a good target in latent phase infection. Thousands of proteins have been classified, in particular new targets belonging to the mycothiol biosynthesis pathway or already validated targets such as the mycolic acid biosynthesis pathway.

In the second part, we studied by means of bioinformatics and computational chemistry the Cyclopropane Mycolic Acid Synthases (CMAS) protein family which belong to the mycolic acid biosynthesis pathway and have been validated as good therapeutic target. These enzymes have S-adenosyl-L-methionine (SAM) dependent methyltransferase activity with a peculiarity, each one of them has a strong selectivity and a specific product, being able to produce ciclopropanes or methyl-alcohol groups from an olefin group. We describe the molecular determinants of the different products in CMAS proposing some testable predictions by doing mutagenesis.

In the last part of this thesis we devoted to design a high throughput docking protocol tailored to the bicarbonate bearing CMAS subfamily (cmaA1-2, pcaA, mmaA4 and umaA) using umaA as reference. We propose different bicarbonate-mimicking groups (Urea, glycine, Carba-

mate, Metoxiacetic acid and Hydroxifuranone) which could act as pharmacophores searching for compounds that have them in the ZINC database. A 30.000 compound database was used to screen against umaA with a posterior free energy of binding estimation done with molecular dynamics and MM-PBSA. We conclude that the best compounds are based on the urea and glycine motifs due to, partly, the hydrogen bond interaction between the amino group in these compounds and a glutamic acid residue of the protein present in the CMAS subgroup.

Summing up, we have developed a bioinformatics pipeline capable of target selection with user-supplied criteria, tested CMAS reaction mechanisms and used the derived information to develop new compounds with bactericidal activity.

Índice

Capítulo 1

Introducción

La búsqueda de nuevos fármacos es una área de intensa investigación. En 2015, la *Food and Drug Administration* de Estados Unidos aprobó 45 nuevos fármacos, de los cuales 1 sólo se trata de un antimicrobiano. (Mullard, 2016) Este número es preocupante dada la incipiente aparición de cepas resistentes y multiresistentes a antibióticos de primera y segunda generación en bacterias como *Staphylococcus aureus, Enterococcus spp., Pseudomonas aeruginosa, Acinetobacter spp.* y *Mycobacterium tuberculosis*. Uno de los puntos críticos en el proceso de descubrimiento de nuevos fármacos trata de la elección y validación del blanco molecular. Dada su relevancia, en este trabajo de tesis, proponemos utilizar herramientas bioinformáticas como el análisis estructural de proteomas, el modelado por homología y la clasificación de proteínas en dominios combinando esta información con resultados experimentales de expresión en diversas condiciones para la elección de blancos moleculares en *Mycobacterium tuberculosis*. A su vez con la ayuda de el *docking* y la dinámica molecular pretendemos hallar y proponer compuestos que puedan actuar como bactericidas particularmente para la fase de latencia del bacilo. Adicionalmente, buscamos proponer una serie de compuestos que puedan ser utilizados como punto de partida para el desarrollo de inhibidores de los blancos moleculares elegidos que en una etapa posterior sean probados *in vitro* e *in vivo*. A lo largo de este capítulo daré una breve introducción a las principales herramientas a utilizar y al sistema biológico elegido.

1.1 Bioinformática

La bioinformática tiene su nacimiento a principios de la década de 1960 con la aparición de las primeras secuencias proteicas y las primeras estructuras cristalográficas (Hagen, 2000), debido a la necesidad de analizar la gran cantidad de datos que se estaban generando en ese entonces. Las primeras aplicaciones bioinformáticas surgen del trabajo del grupo de Margaret Dayhoff realizado para ensamblar una secuencia proteica a partir de los péptidos derivados de la secuenciación de Edman (Margaret Oakley Dayhoff, & Ledley, 1962; Margaret O Dayhoff, 1965). A partir de allí la explosión de datos provocada por el descubrimiento del código genético (Crick et al., 1961) en la década de 1960 y aparición de las técnicas de secuenciación de ADN (Sanger, Nicklen, & Coulson, 1977) en la década de 1970 provocaron la necesidad del desarrollo de técnicas computacionales para el manejo y la interpretación de los crecientes datos biológicos. Las principales técnicas de la bioinformática clásica se enmarcan en el problema del alineamiento entre secuencias, sean tanto de ácidos nucleicos (ADN y ARN) como de proteínas. Estas técnicas permitieron realizar los primeros estudios filogenéticos a escala molecular (la evolución molecular), como así también comprender la relevancia funcional de los aminoácidos en las proteínas a partir de la construcción de familias de proteínas como PFAM (Bateman et al., 2004), PROSITE (Hulo et al., 2006) y SMART (Letunic, Doerks, & Bork, 2012). El otro hecho relevante para el surgimiento de la bioinformática son las bases de datos en donde los resultados están disponibles. El principal repositorio de bases de datos biológicas es NCBI (National Center for Biotechnology Informatiomn) dependiente de NIH-USA creado en 1988. En NCBI se encuentran disponibles unas 67 bases de datos. Ejemplos de estas son GenBank, en donde se depositan secuencias de ADN; GEO - *Gene Expression Omnibus*, una base de datos de expresión de genes y PubMed, una base de datos de publicaciones científicas relacionadas con biomedicina. En la figura 1.1 se muestra la cantidad de pares de bases depositados en el NCBI junto con el número de usuarios que la utilizan, se observa una explosión de datos depositados a partir de la publicación del primer borrador del genoma humano en 1999.

Por otro lado, si bien no sufrió la misma explosión en información presentada por el NCBI, principalmente debido a la dificultad que presenta obtener la información experimental, la base de datos de estructuras proteicas PDB (Protein Data Bank) (Bank, 1971) presenta actualmente 115.000 estructuras disponibles (Ver Figura 1.2). Si bien este número cubre un espacio relativa-

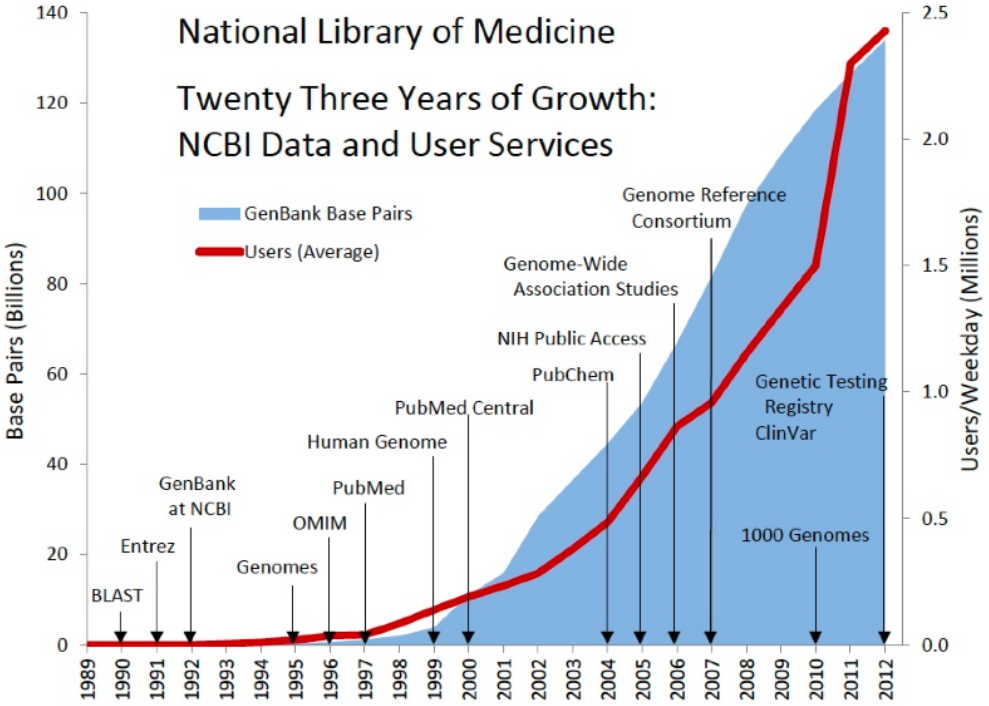


Figura 1.1: **Número de pares de bases (en miles de millones) depositados en el NCBI** (National Center for Biotecnology Information), NIH, USA, en azul. En rojo se indica el numero de usuarios por día. Las flechas indican la aparición de herramientas o bases de datos a lo largo de los años, como BLAST o OMIM. Gráfico del National Library of Medicine.

mente pequeño de las proteínas en la naturaleza, a partir de técnicas de modelado comparativo, explicadas más adelante en esta introducción, es posible expandir ese número en órdenes de magnitud, ya que la conservación estructural es mucho mayor que la conservación de secuencia. (Illergård, Ardell, & Elofsson, 2009)

1.1.1 Drogabilidad

La drogabilidad es un concepto utilizado para describir que tan relevante biológicamente y que tan factible es inhibir una proteína de un organismo en particular (Hopkins, & Groom, 2002; Cheng et al., 2007). Desde un punto de vista puramente estructural, está relacionado con la probabilidad que una molécula pequeña se una a una proteína con una alta afinidad (<1 micromolar), este concepto también se encuentra descrito en la literatura científica como *'bindability'* y a la proteína la denominamos como *bindable*

Los primeros reportes de la drogabilidad del genoma de un organismo, basados en el número de blancos pertenecientes a dominios que se saben drogables, arrojó valores entre 10 y 14 % para el genoma humano. (Hopkins, & Groom, 2002) Las proteínas drogables deben tener bolsil-

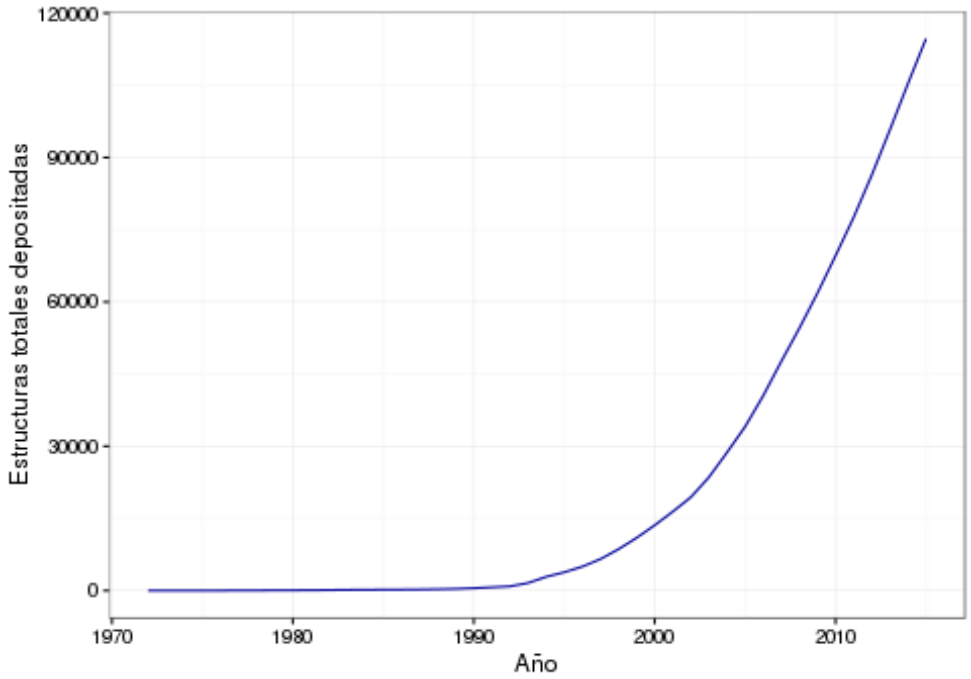


Figura 1.2: **Numero acumulado total de estructuras depositadas en el PDB por año.** Elaborado en base a datos del RSCB

los con propiedades adecuadas que permitan la unión de compuestos tipo droga , moléculas que pueden ser ingeridas por via oral y que modulan un proceso biológico de interés. (Davis et al., 2007; Kinnings et al., 2010) Recientemente, el grupo de los Dres. Barril y Luque de la Universitat de Barcelona desarrolló un método rápido para la predicción de drogabilidad basado en un algoritmo de detección de bolsillos de código abierto denominado fpocket. (Cheng et al., 2007; Xie, & Bourne, 2007; Halgren, 2009) Este método combina varios descriptores fisicoquímicos para estimar la drogabilidad de cada bolsillo proteico y puede ser utilizado a escala genómica dada su rapidez. (Schmidtke, & Barril, 2010).

En el desarrollo de esta tesis se utilizó la capacidad de fpocket de determinar si un bolsillo es capaz de unir un compuesto tipo droga. Pero, esto no basta para determinar si en verdad el blanco seleccionado es drogable, también hay que estimar si la unión de un compuesto modulará la actividad como se desea y si finalmente producirá el efecto deseado. En particular, para la tuberculosis, organismo con el cual trabajamos en esta tesis, dadas las peculiaridades del bacilo es necesario tener en cuenta también el medio en el cual vive durante la fase de latencia y las condiciones que afronta (hipoxia, estrés derivado de especies reactivas de oxígeno y nitrógeno -ERON-, falta de nutrientes).

En esta tesis decidimos diseñar una función que dependiendo de diferentes propiedades

de los genes les asigne un valor de puntuación que nos permita clasificarlo. Como nuestro objetivo incluye decidir los genes relevantes en la fase de latencia usamos para la puntuación información de expresión contextual experimental tomada de microarreglos de sobreexpresión de genes en condiciones que imitan el proceso infectivo. Por último esta función toma en cuenta la importancia que tiene una dada proteína en en el contexto de una vía metabólica, denominado drogabilidad contextual. Esto permite elegir los blancos basados no solo en las propiedades estructurales sino también en la relevancia biológica que tiene para el organismo en cuestión siendo esenciales para el organismo y produciendo un efecto bactericida.

1.1.2 Bioinformática estructural

El Proteín Data Bank (PDB) fue fundado en el año 1971 por el Dr. Walter Hamilton en Brookhaven National Laboratory conteniendo solo 7 estructuras cristalográficas. Actualmente es el repositorio mundial de 115.000 estructuras tridimensionales de macromoléculas grandes como son las proteínas, ácidos nucleicos y sus complejos con lípidos y azucares como así también con diversos ligandos, entre ellos compuestos tipo fármaco. Aproximadamente el 90 % de las estructuras depositadas en el PDB provienen de experimentos de difracción de rayos-X. Dado que la obtención de cristales y obtención de estructuras a partir de los experimentos de difracción resulta complicado en muchos casos se puede valer de la bioinformática estructural para computar modelos de proteínas de interés a partir de las estructuras depositadas en el PDB. Este procedimiento se denomina modelado por homología o modelado comparativo. El software por excelencia utilizado para dicho procedimiento se llama MODELLER (Eswar et al., 2008), aunque existen otras opciones basados en los mismos conceptos como Swiss-Model (Biasini et al., 2014) y Phyre2 (Kelley et al., 2015).

MODELLER necesita de un alineamiento de secuencia entre la proteína objetivo (*target*) y el molde (*template*). Utilizando ese alineamiento como base y generando una serie de restricciones espaciales en la cadena principal (los carbonos alfa) genera un modelo de la proteína objetivo que luego refina utilizando los parámetros del campo de fuerza de CHARMM. En el caso que no haya coincidencia entre el molde y el objetivo utiliza una función de modelado *ab initio* para completar los huecos. La calidad del modelo depende principalmente de la identidad de secuencia (el porcentaje de coincidencia entre las dos secuencias)y la cobertura (que

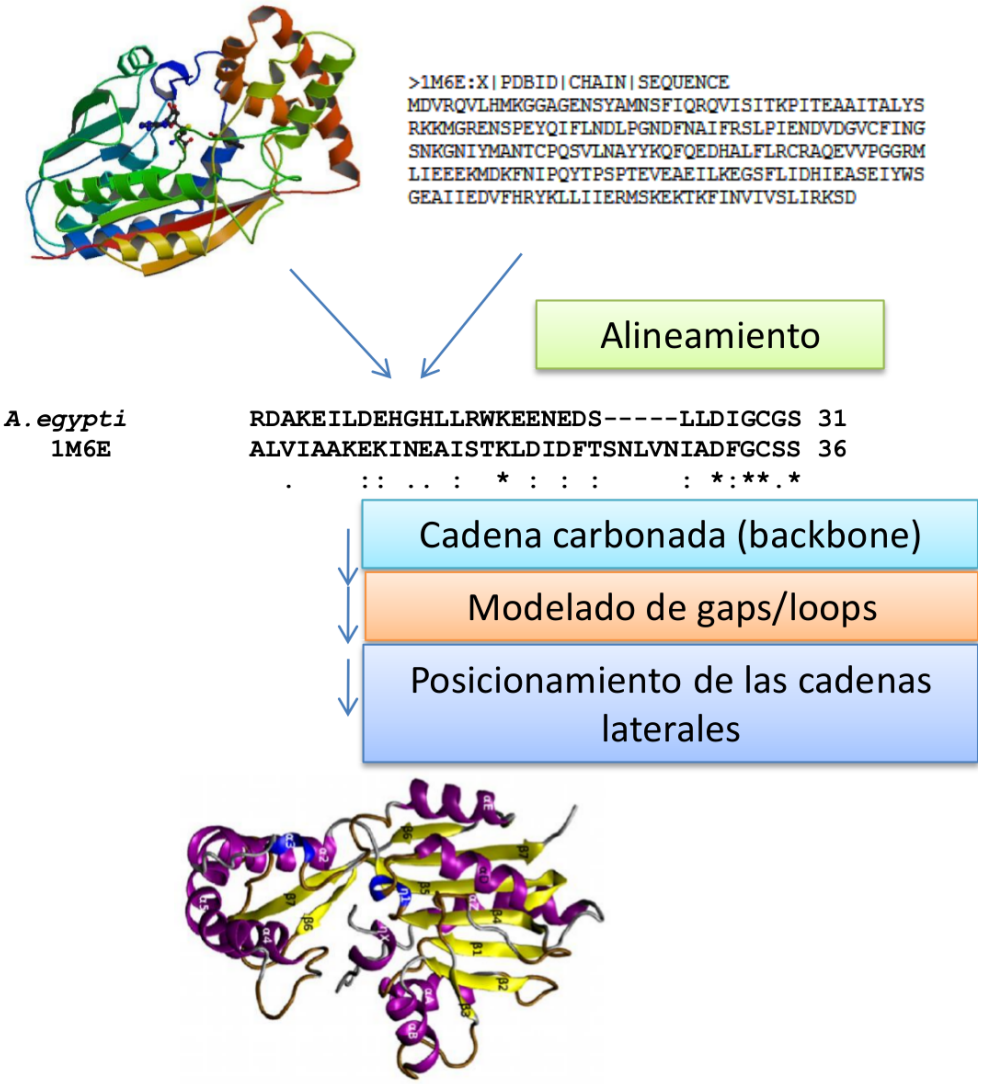


Figura 1.3: **Procedimiento de modelado por homología.** El modelado requiere de una estructura que actúa como molde y la secuencia de ambas, que debe ser alineada tomando en cuenta la presencia de información estructural en el molde. El modelado se realiza por partes, tomando en primer lugar las restricciones derivadas de posicionar la cadena principal (representada como los carbonos α), luego se posiciona la cadena principal de regiones de las cuales no se dispone de información, en general flexibles, mediante técnicas *ab initio*. Por último, se insertan las cadenas laterales y se realiza una optimización global del modelo.

porción de secuencia fue posible alinear, sin importar si existe coincidencia) del alineamiento entre molde y objetivo como así también que los sitios importantes funcionalmente se encuentren correctamente alineados. Con este tipo de procedimientos es posible obtener de las 58.000 estructuras no redundantes depositadas en el PDB unos 34 millones de modelos comparativos generados por MODELLER de manera totalmente automática. Estos resultados se encuentran depositados en la base de datos ModBase (Pieper et al., 2011). Un paso fundamental para la correcta elección de moldes para realizar modelado por homología es tener a las proteínas clasificadas en dominios o familias que compartan relaciones evolutivas (Estructura-Función).

1.1.3 Familias de proteínas

Existe un interés en poder agrupar a las proteínas en conjuntos que compartan características funcionales, estructurales y filogenéticas. Este interés radica en la posibilidad de estudiar un número pequeño de sistemas modelo y luego extrapolar características relevantes como son la estructura y función a un sinnúmero de otras proteínas aún no estudiadas, proceso coloquialmente denominado como de culpa por asociación. Estos conjuntos proteicos se los conoce como Dominios o Familias de proteínas.

Existen diversas formas de construir una definición de dominio proteico, la que mayormente se utiliza es la derivada de modelos ocultos de Markov (Hidden Markov Model, HMM) dado que tiene una base matemática sólida. Brevemente, los modelos ocultos de Markov son una técnica matemático-estadística para representar series temporales, en el caso particular de los modelos ocultos solamente se conoce los observables y no el estado particular en el que uno se encuentra. En el caso de los HMM de proteínas la secuencia se modela como una sucesión (serie temporal) de estados (ocultos) que emiten aminoácidos. A partir de un alineamiento múltiple de secuencias de proteínas que pertenecen a una misma familia se puede estimar los parámetros necesarios para crear un HMM específico para esa familia.

Para clasificar las proteínas en Dominios el HMM, debe ser capaz de resolver tres problemas. El primero es como determinar si una secuencia puede provenir de un HMM particular, es decir, si pertenece a ese conjunto, denominado problema de la puntuación. El segundo problema es el denominado del alineamiento que implica saber cual es el alineamiento múltiple que maximiza el puntaje de todos los miembros de la familia. El último es como a partir de un alineamiento múltiple de secuencia se computan las probabilidades necesarias para poder definir este dominio de forma inequívoca, denominado problema del entrenamiento. Estos últimos dos problemas se resuelven de manera iterativa. Las principales ventajas que tienen los HMM de proteínas están en su velocidad, dado que no se requiere recurrir a alineamientos múltiples una vez construido el modelo, no asume puntajes *a priori* y tienen en cuenta de forma intrínseca las inserciones/deleciones.

Existen diversas implementaciones de HMM como método de clasificación de proteínas las más conocidas son, Pfam (Bateman et al., 2004),, SMART (Schultz et al., 1998), TIGRFAM (Haft, Selengut, & White, 2003), PIRSF (Wu et al., 2004), PANTHER (Thomas et al., 2003),

Superfamily (Gough et al., 2001) and Gene3D (Yeats et al., 2006).

La base de datos de dominios más utilizada es Pfam (Bateman et al., 2004), derivada inicialmente de la base de datos curada Swiss-Prot. Combina un sistema de alineamiento automático y una curación manual para la generación de las familias proteicas. Actualmente posee 16200 familias. La clasificación de proteínas en dominios permite realizar transferencia de estructura y/o función de miembros estudiados en detalle a otros que no han sido estudiados con tanto detalle. El grado de certeza dependerá de que tan estrecha sea la relación evolutiva entre ambas. Esta clasificación puede ayudar de manera rápida a seleccionar las proteínas de un organismo pudiendo enfocarse en las que se tenga más certeza de su función y rol biológico, fundamental para cualquier proceso de búsqueda de blancos moleculares. Otra propiedad interesante que ha sido probada en el pasado es la capacidad de proteínas pertenecientes a un mismo dominio a unir compuestos con una afinidad diferencial. Otro caso en el cual resulta de utilidad la clasificación en dominios es a la hora de realizar un modelado por homología. A través de HMMs el servidor PHYRE2 (Kelley et al., 2015) utiliza la información evolutiva que existe en un HMM de proteínas para facilitar la elección de moldes para realizar el modelado por homología.

Tener a las proteínas clasificadas en familias permite realizar de manera más precisa el proceso de modelado comparativo como así también facilita enormemente comprender el rol de ciertas porciones de la misma, por ejemplo, la ubicación del sitio activo si se trata de una enzima, para miembros no tan estudiados. Las estructuras (sean tanto experimentales o modelos, si la calidad lo permite) pueden ser utilizadas como punto de partida para realizar una búsqueda virtual de compuestos que funcionen como inhibidores.

1.2 Desarrollo de fármacos

El desarrollo de fármacos tiene distintas etapas, que están descritas en la figura 1.4. Dichas etapas son: i) La identificación del blanco molecular que se desea atacar; ii) la realización de una búsqueda mediante técnicas masivas *in vitro* o *in silico* de compuestos líderes, capaces de modular (inhibir) la actividad del blanco; iii) la validación del blanco y la optimización del fármaco para pasar a un compuesto líder como así también mejorar la seguridad medicinal del mismo (evitar efectos de *off-target*). Estas etapas se encuentran en lo que se denomina la fase pre-clínica. Las otras etapas, de los ensayos clínicos propiamente dichos se denominan: Fase I,

Fase II, Fase III y Fase IV. En la primer fase se realizan las pruebas de seguridad del fármaco en pacientes sanos. En las dos subsiguientes etapas se sigue evaluando la seguridad del fármaco pero a su vez se evalúa a eficacia del mismo como así también la dosis, incrementando el tamaño de la población entre la fase II y la fase III. La última etapa es la fase de farmacovigilancia (Fase IV), en donde el fármaco se encuentra aprobado pero se sigue evaluando su seguridad y eficacia en una población aún mayor.

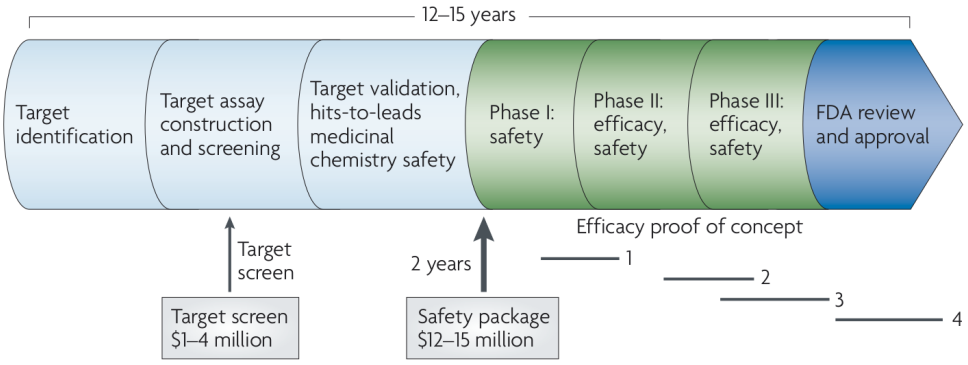


Figura 1.4: **Esquema de trabajo para el desarrollo de fármacos.** El esquema se puede dividir en dos etapas, una pre-clínica y una clínica. En la primer etapa se realizan todos los ensayos de búsqueda, validación, optimización y seguridad de las moléculas encontradas. En la segunda etapa se realizan los ensayos clínicos de efectividad y seguridad con poblaciones cada vez más grandes y, una vez aprobado, se pasa a la etapa de farmacovigilancia. Tomado de (Roses, 2008)

Las estrategias utilizadas en esta tesis se concentran enmarcadas en las etapas pre-clínicas, por esa razón se hará un comentario más detallado de estas etapas. La primera, y por cierto la más importante, es la validación del blanco molecular a atacar. En otras palabras, que la proteína para la cual deseamos desarrollar un fármaco esta involucrada en el proceso patológico que deseamos tratar. Por ejemplo, las estatinas, son drogas exitosas que actúan bajando los niveles de colesterol cuyo blanco molecular es la enzima HMG-CoA reductasa, punto de control de la vía de síntesis de colesterol endógeno *de novo* (Branchi et al., 1995). Esta etapa es donde la mayor cantidad de fármacos falla y su inadecuada elección hace al éxito futuro de cualquiera de las etapas subsiguientes, dado que si el mecanismo molecular no es conocido o no se comprende bien que ocurre al inhibir la función de una enzima a nivel sistémico puede tener consecuencias inesperadas. La falta de un correcto desarrollo de validaciones biológicas ha provocado el aumento de la tasa de *attrition* (compuestos que no llegan a la clínica), principalmente en desarrollos oncológicos. (Hutchinson, & Kirk, 2011; Kola, & Landis, 2004).

La elección y prueba de la biblioteca de compuestos es la siguiente etapa. Esta etapa puede

se abordada por metodologías experimentales o un enfoque mixto experimental-simulación. Desde un punto de vista meramente experimental, se enfrentan células o proteínas recombinantes a la biblioteca de compuestos, en un diseño que permita evaluar fácil y de manera automática la actividad proteica o celular (Sundberg, 2000; Bleicher et al., 2003; Pausch, 1997; Collins, Torrero, & Franzblau, 1998; Ananthan et al., 2009) Generalmente se prueban desde cientos de miles a millones de compuestos lo que hace a esta estrategia abordable sólo por las grandes farmacéuticas o centros públicos de gran envergadura como el *Chemical Genomics Center* de NIH en Estados Unidos. Por lo cual, en la mayoría de los casos, las estrategias computacionales se utilizan como un filtro grueso para acotar la búsqueda a una familia pequeña de compuestos, con los cuales se puede realizar las pruebas *in vitro* correspondientes.

El caso más conocido de éxito de las metodologías computacionales para encontrar nuevos compuestos tipo-droga es el de la HIV-proteasa.(Wlodawer, & Vondrasek, 1998; Lam et al., 1994). Por otro lado, se evidencia un resurgimiento de las estrategias de búsqueda de fármacos basadas en fenotipo (*Phenotypic-based screening*) (Mullard, 2015). En este tipo de técnicas se busca, primero, encontrar compuestos bioactivos al observar un cambio fenotípico y registrarlo como por ejemplo el nivel de expresión de un gen acoplado a una proteína fluorescente. Luego es necesario encontrar la molécula blanco para poder elucidar el mecanismo de acción. En general para encontrar los blancos moleculares se requiere generar mutantes resistentes al compuestos hallado y realizar una secuenciación del genoma entero para proponer posibles blancos. Este tipo de experimentos es mucho más factible dada la baja en los costos de las técnicas masivas de secuenciación, que permiten encontrar todas las posibles mutaciones que causarían la resistencia.

Una vez encontrada una serie de compuestos activos por técnicas masivas, se pasa a la etapa de validación del blanco molecular y optimización de compuestos. En general esta etapa involucra la mejora de las propiedades de unión/inhibición del compuesto en cuestión (la farmacodinámica) y la mejora de sus propiedades farmacocinéticas (las denominadas propiedades ADMET, absorción, distribución, metabolismo, excreción y toxicología). En esta etapa también se realizan usualmente las primeras pruebas en animales.

En el contexto del proceso mencionado para la búsqueda de nuevos fármacos, en esta tesis, se realizó una búsqueda de posibles blancos contra la fase latente de *Mtb.* teniendo especial

cuidado de que sean biológicamente relevantes y su disrupción cause un efecto bactericida. En segundo lugar, para contribuir a la búsqueda de compuestos líderes se realizó una búsqueda virtual de compuestos para una familia de proteínas particular, las *Cyclopropane Mycolic acid Synthases*, que forman parte de la vía de síntesis de ácido micólico, esencial para el bacilo.

1.3 Métodos computacionales para el desarrollo de fármacos

1.3.1 Métodos basados en las propiedades de los ligandos

Los primeros métodos para establecer una relación cuantitativa entre la estructura de una molécula y su actividad fueron desarrollados por Fujita y Hansch a mediados de la década de 1960 y se basaban en la simple premisa de que compuestos similares en estructura tendrían actividades similares (relaciones de estructura-función o SAR en inglés). Al conjunto de reglas y procedimientos que establecen una relación entre las propiedades físicoquímicas, derivadas de la estructura de un compuesto, con su actividad biológica, determinada por un ensayo estandarizado de forma cuantitativa, se lo denomina QSAR (Quantitative structure-activity relationship) (Fujita, & Hansch, 1967). La finalidad del QSAR es construir un modelo que permita predecir si un compuesto nuevo tendrá actividad, será selectivo o si tendrá éxito como fármaco. Para poder llevar adelante este tipo de estudios, es fundamental contar con información precisa y detallada de un gran número de compuestos diferentes (decenas o centenas) con el fin de establecer un modelo de predicción confiable. El modelo depende principalmente de la calidad de los datos biológicos, la elección de descriptores físicoquímicos (grupos funcionales o farmacóforos) y métodos estadísticos empleados para el análisis de la información. Partiendo del hecho de que un farmacóforo es un conjunto de rasgos estéricos y electrónicos que son necesarios para asegurar interacciones supramoleculares con un blanco biológico específico, en la química computacional moderna, estos grupos funcionales se emplean para definir los rasgos esenciales de una o más moléculas con la misma actividad biológica. Una base de datos de diversos compuestos químicos puede ser analizada en búsqueda de moléculas que exhiben las mismas características y que por ello se diferencian del resto. Estos modelos tienen un valor predictivo medio dependiendo principalmente de la calidad de los ensayos bioquímicos disponibles y la diversidad química de las moléculas utilizadas para entrenar el modelo. Un

excelente caso de estudio para estos métodos son las proteína quinasas en donde existen gran cantidad datos disponibles acerca de la afinidad de un sinnúmero de moléculas en donde los métodos de QSAR se lucen. (Falchi et al., 2009; D. Huang et al., 2012; Sprous et al., 2006) Incluso se ha llegado a utilizar técnicas de aprendizaje automático (*machine learning* por su nombre en inglés) para descubrir nuevas moléculas con perfiles polifarmacológicos, es decir que tienen como blanco a más de una proteína. (Besnard et al., 2012). La principal desventaja de este tipo de metodologías basadas en la información de los ligandos es que se requiere de una gran cantidad de datos experimentales *a priori* para poder derivar un modelo. Esta información nunca está disponible, en el caso de tratarse de un blanco novedoso, como aquellos que son objetivo de esta tesis.

1.3.2 Métodos basados en las propiedades de los blancos

En general, las estrategias de *docking* (encastre, nos tomaremos la licencia de utilizar el vocablo en inglés) tienen como objetivo predecir la estructura del complejo proteína-ligando en base a la estructura del receptor y del ligando separados. Las técnicas comúnmente utilizadas para realizar *docking* automático se pueden agrupar principalmente en dos categorías: métodos de coincidencia geométrica y métodos de simulación de formación de complejos. Los primeros crean un modelo del sitio activo, donde típicamente se incluyen zonas de interacción de puente hidrógeno y zonas que son estéricamente accesibles. Con este modelo intentan ubicar un determinado inhibidor en una determinada conformación, considerándolo usualmente como un cuerpo rígido. Este es el caso del programa DOCK, el cuál es muy eficiente para hacer un barrido rápido bibliotecas de compuestos (Lang et al., 2009) utilizando la coincidencia geométrica. La segunda clase de métodos de *docking* se corresponde con aquellos que describen con mayor grado de detalle la estructura del ligando y del receptor, ubicando el ligando de manera aleatoria fuera de la proteína para luego explorar traslaciones, orientaciones y conformaciones hasta que el estado ideal del complejo proteína-ligando sea encontrado. Esta técnica es usualmente más lentas que los métodos de coincidencia geométrica, pero permiten que la flexibilidad del ligando sea considerada. En este contexto, programas como AUTODOCK (Morris et al., 2009) y rDock (Ruiz-Carmona et al., 2014) implementan búsquedas basadas en métodos rápidos de evaluación de la energía de interacción proteína-ligando y búsquedas basadas en algoritmos

genéticos lamarkianos permitiendo estimar el modo de unión de manera aceptable.

La utilización de las técnicas de *docking* para evaluar el modo de unión de millones de compuestos y determinar cuáles de ellos poseen potencialmente una alta afinidad al receptor se denomina búsqueda virtual o *Virtual Screening* (VS). El VS es posible gracias a la ganancia en velocidad por la simplificación del sistema. Este hecho permite probar millones de compuestos en unas semanas en un *cluster* de tamaño moderado (200 procesadores) filtrando rápidamente los que no tienen actividad y priorizando aquellos que si la tienen. Este es el principal mérito de las metodologías de *docking*, reducir de millones a unos cientos los posibles compuestos con actividad biológica.

Las herramientas presentadas en esta sección (Bases de datos biológicas y estructurales, clasificación de proteínas en familias/dominios, drogabilidad, *docking*) serán utilizadas para armar una base de datos de un organismo particular de interés clínico como *Mycobacterium tuberculosis* en pos de clasificar y proponer nuevos blancos terapéuticos y los respectivos compuestos que sirvan como punto de partida para inhibirlos.

1.4 *Mycobacterium tuberculosis*

De acuerdo con el último reporte global de la tuberculosis (TB) de la Organización Mundial de la Salud (OMS), en 2014 un estimado de 8.6 millones de personas en el mundo desarrollaron la enfermedad, llevando a 1.3 millones de muertes (incluyendo 320.000 infectadas con el virus del HIV). (WHO, 2014) (Ver figura 1.5) Las terapias tradicionales para TB involucran un largo tratamiento con drogas de primera generación como son la isoniazida, la rifampicina, la pirazinamida y el etambutol. (Caminero et al., 2010)

Sin embargo, la aparición de cepas multiresistentes (MDR) y extremadamente resistentes (XDR) de *Mycobacterium tuberculosis (Mtb)* (Figura 1.6) y la interacción negativa entre varias de las drogas arriba descriptas y las utilizadas para tratar el SIDA (y algunas otras enfermedades) han demostrado la urgente necesidad de desarrollo de nuevas terapias contra Mtb. (Koul, Arnoult, et al., 2011; Russell, Barry, & Flynn, 2010) En la última década la secuenciación del genoma de *Mtb* y las técnicas bioinformáticas para manejar preguntas en esa escala, que comprende aproximadamente 4.000 genes, ha abierto la posibilidad de hallar nuevas estrategias terapéuticas.(Lew et al., 2011; Agüero et al., 2008; Schilling et al., 1999; Reddy et al.,

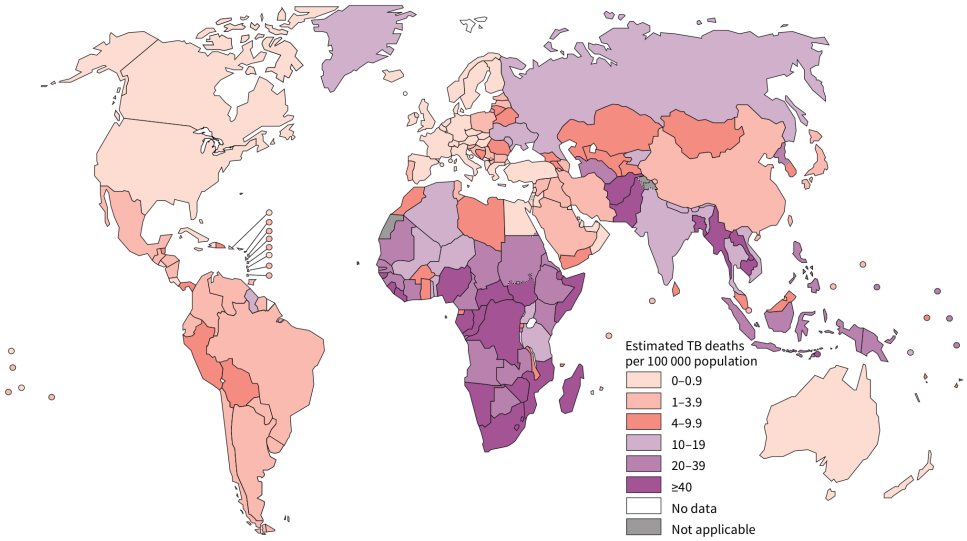


Figura 1.5: **Cantidad de muertes de TB a nivel global en cientos de miles.** Tomado de (WHO, 2014)

2009)

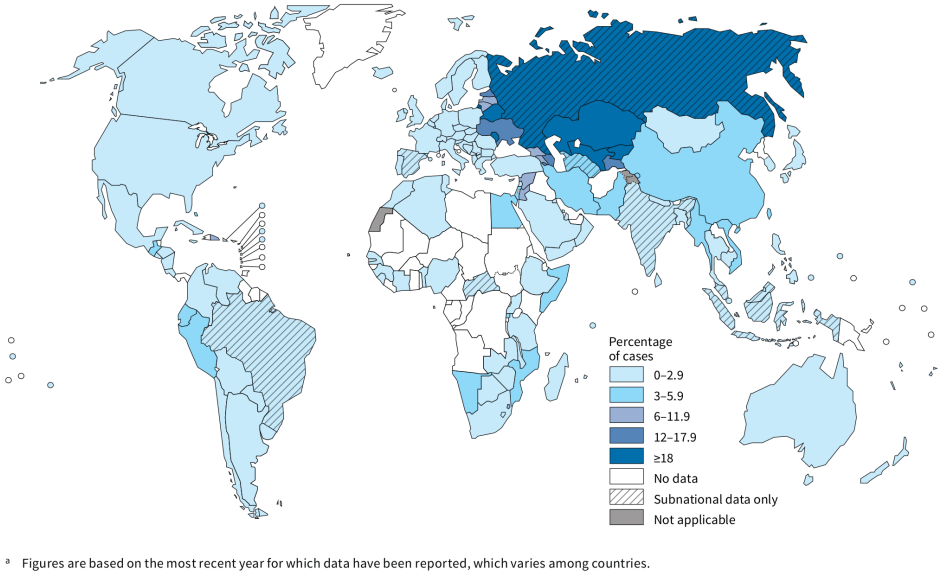


Figura 1.6: **Porcentaje de casos de MDR TB sobre el total de casos.** Tomado de (WHO, 2014)

En particular, el análisis del genoma tiene el potencial de permitir obtener información valiosa para el desarrollo de nuevas terapias e intervenciones para tratar la enfermedad. En años recientes varias bases de datos que integran datos genómicos de *Mtb.* de diversa índole como Tuberculist, tbvar, TBDB, o TDR-Targets han aparecido. (Lew et al., 2011; Agüero et al., 2008; Schilling et al., 1999; Reddy et al., 2009) Bases de datos como Tuberculist, tbvar y TBDB, se centran en información genómica curada y de expresión de genes únicamente. TDR-Targets,

si bien es una plataforma pensada para elegir blancos moleculares, no cuenta con información de *bindability* de los blancos ni información contextual de los genes. Estrategias similares fueron utilizadas para identificar potenciales blancos de drogas en *Mtb.* pero ninguno realizó un ensayo con todo el proteoma. Por esta razón realizaremos en esta tesis un análisis abarcando todo el proteoma de *Mtb.* incluyendo información contextual (expresión de genes en diversas condiciones relevantes para la infección).

1.4.1 El tratamiento actual contra TB

Los compuestos utilizados actualmente para el tratamiento de la tuberculosis son: la izoniazida, cuyo blanco molecular es la Enoil-ACP reductasa (inhA) una enzima que forma parte de la vía de síntesis de ácido micólico; la rifampicina, cuyo blanco molecular es la RNA Polimierasa B (rpoB); la estreptomicina, atacando la actividad de la proteína ribosomal S12; el etanbutol, cuyo blanco molecular es el complejo responsable de polimerizar la arabinosa en arabinan que conforma los arabinogalactanos de la pared de la mycobacteria; la pirazinamida, cuyo blanco es el complejo de síntesis de ácidos grasos FAS I y la trans-traducción mediada por la proteína ribosomal S1; la ofloxacina, que actúa inhibiendo la actividad de la ADN Girasa B; cicloserina, que actúa inhibiendo la síntesis de péptidoglicano a través de su acción con la Alanina racemasa y con la D-alanila ligasa; la kanamicina, cuyo mecanismo de acción esta ligado a su unión al ribosoma 30S inhibiendo la síntesis proteica y la capreomicina, cuyo mecanismo de acción es desconocido.

El tratamiento con drogas de primera linea involucra tomar durante 6 meses isoniazida, rifampicina, pirazinamida, etanbutol y streptomicina. (Organization, 2010) Si existen evidencias de cepas multiresistentes se pasa a un régimen en donde se administra cicloserina y alguna fluoroquinolona (ofloxacina). El problema radica es que estos agentes tienen un sin número de contraindicaciones entre las que se encuentran gastritis, efectos dermatológicos, neuropatías periféricas, depresión y ansiedad. (Furin et al., 2001) Estos efectos adversos se incrementan si existe una co infección con HIV, siendo aún más delicado el tratamiento. A pesar de que recientemente, luego de 40 años, la FDA ha aprobado una nueva droga para tratar las infecciones con cepas MDR, la *bedaquiline* su modo de acción involucra la inhibición de la la ATP sintetasa (Koul, Vranckx, et al., 2014) lo que genera severos efectos adversos. Por otro lado, como se de-

scribirá más adelante, no existen drogas capaces de atacar al bacilo dentro del macrófago cuando se encuentra en el estado de latencia. Dada la complejidad de la estrategia terapéutica utilizada, es fundamental encontrar nuevos fármacos con capacidad bactericida contra *Mycobacterium tuberculosis*.

1.4.2 El ácido micolico y las *Cyclopropane Mycolic Acid Synthases* (CMAS)

Mycobacterium tuberculosis posee una pared externa casi impermeable que lo protege de las condiciones hostiles que enfrenta dentro del macrófago (como son por ejemplo, el ataque con especies reactivas oxígeno y nitrógeno, ERON). (Abdallah et al., 2007). Esta pared externa está formada por glicolípidos (Marrakchi, Lanéelle, & Daffé, 2014), un ester de un disacárido (trealosa) con un ácido micólico (AM, un ácido graso de cadena muy larga, con 40 a 60 átomos de carbono). El espesor de dicha pared le brinda al bacilo propiedades interesantes, como una gran impermeabilidad, mientras que algunos de los componentes de la pared externa son responsables de modular la respuesta inmune del huésped. (Marrakchi, Lanéelle, & Daffé, 2014; D. Barkan et al., 2012) (ver Figura 1.7) En este sentido resulta atractivo el desarrollo de fármacos para la inhibición de proteínas de esta vía.

La primer parte de la síntesis *de novo* de ácidos micólicos es compartida con los ácidos grasos (AG) de cadena corta (presente en la mayor parte de las bacterias y eucariotas), el ciclo de síntesis de FAS I (Fatty Acid Synthase I) (Marrakchi, Lanéelle, & Daffé, 2014) generando AG de hasta 16-18 átomos de carbono a partir de la condensación de, primero, acetil-CoA y malonil-CoA, para luego condensar al acilo creciente cadenas de malonil-CoA, ver Figura 1.8. A partir de allí, la elongación de la cadena es responsabilidad del segundo complejo enzimático de síntesis de ácidos grasos, presente en el suborden *Corynebacterineae*, FAS II (Fatty acid synthase II). A diferencia del complejo I, la condensación se da a partir de malonil-ACP, que condensado con el Acil-CoA forma beta-cetoacil-ACP (realizada por la 3-oxoacil-[ACP] sintasa 3, FabH). El ciclo esta compuesto por unas serie de reducciones, llevadas a cabo por la 3-oxoacil-[ACP] reductasa (MabA) y la enoil-[ACP] reductasa dependiente de NADH (InhA), una deshidratación realizadas por el complejo HadAB/HadBC ((3R)-hidroxiacil-ACP dehidratasa). Las posteriores condensaciones son realizadas por el complejo KasA/KasB (3-oxoacil-[ACP] synthase 1 y 2). Luego de la producción de las cadenas largas se produce la introducción de

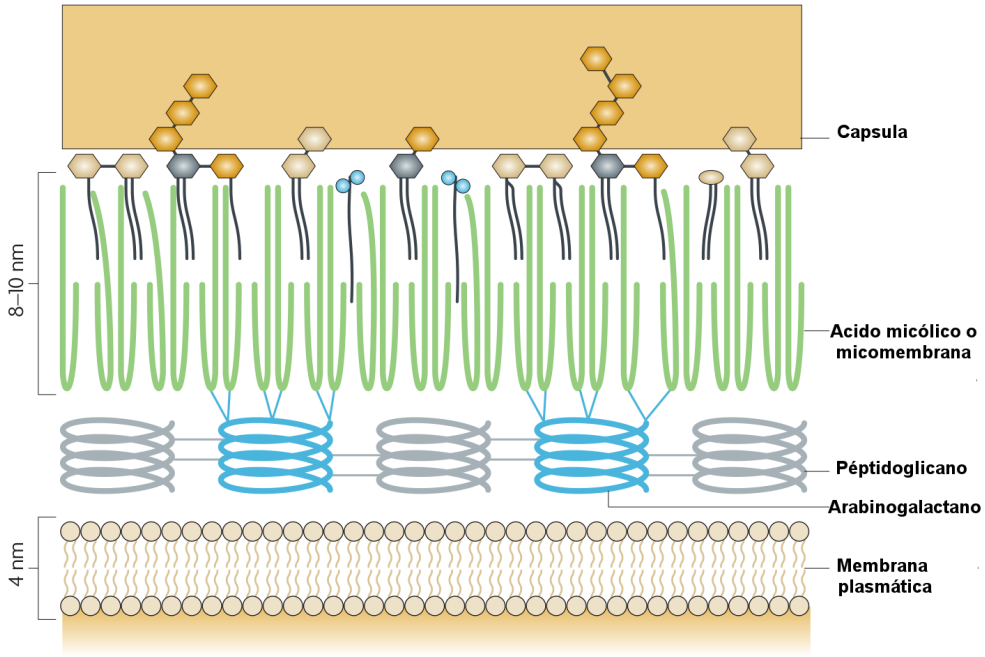


Figura 1.7: **Representación esquemática de la pared de *Mycobacterium tuberculosis*** La pared se encuentra compuesta por tres componentes covalentemente unidos, el ácido micólico (en verde), los peptidoglicanos (en gris) y los arabinogalactanos (en celeste). La parte externa esta compuesta por la capsula (polisacaridos, glucano y arabinomanan) y lípidos libres complejos que se intercalan entre los AM. Adaptada de (Abdallah et al., 2007)

las dos insaturaciones características de los AM realizadas por una o más enzimas no caracterizadas hoy en día. Estos dobles enlaces son modificados finalmente por las actividad de las Cyclopropane Mycolic Acid Synthases (CMAS) para formar grupos ciclopropanos, metil-olefina y metil-éteres que conforman los ácidos micólicos tipo alfa, ceto, hidroxi y metoxi respectivamente (Marrakchi, Lanéelle, & Daffé, 2014). Estas modificaciones son fundamentales para que la pared tenga las propiedades de permeabilidad necesarias para la supervivencia dentro del granuloma. (Marrakchi, Lanéelle, & Daffé, 2014; D. Barkan et al., 2012) El ciclo de síntesis se completa con la activación, condensación con un ácido graso de cadena más corta (24 a 26 átomos de carbono), unión a trealosa y posterior transporte a la pared externa. Dado que se trata de una vía de síntesis única en las mycobacterias y que presenta un rol biológico importante en la infección y respuesta inmune del huésped, la vía de síntesis de ácidos micólicos resulta de gran atractivo para el desarrollo de fármacos. (Marrakchi, Lanéelle, & Daffé, 2014) En este sentido, el fármaco de primera linea, isoniazida inhibe la actividad de InhA y al día de hoy existen un gran número de fármacos en distintas etapas de pruebas que buscan inhibir esta importante ruta metabólica. (Ver figura 1.8)

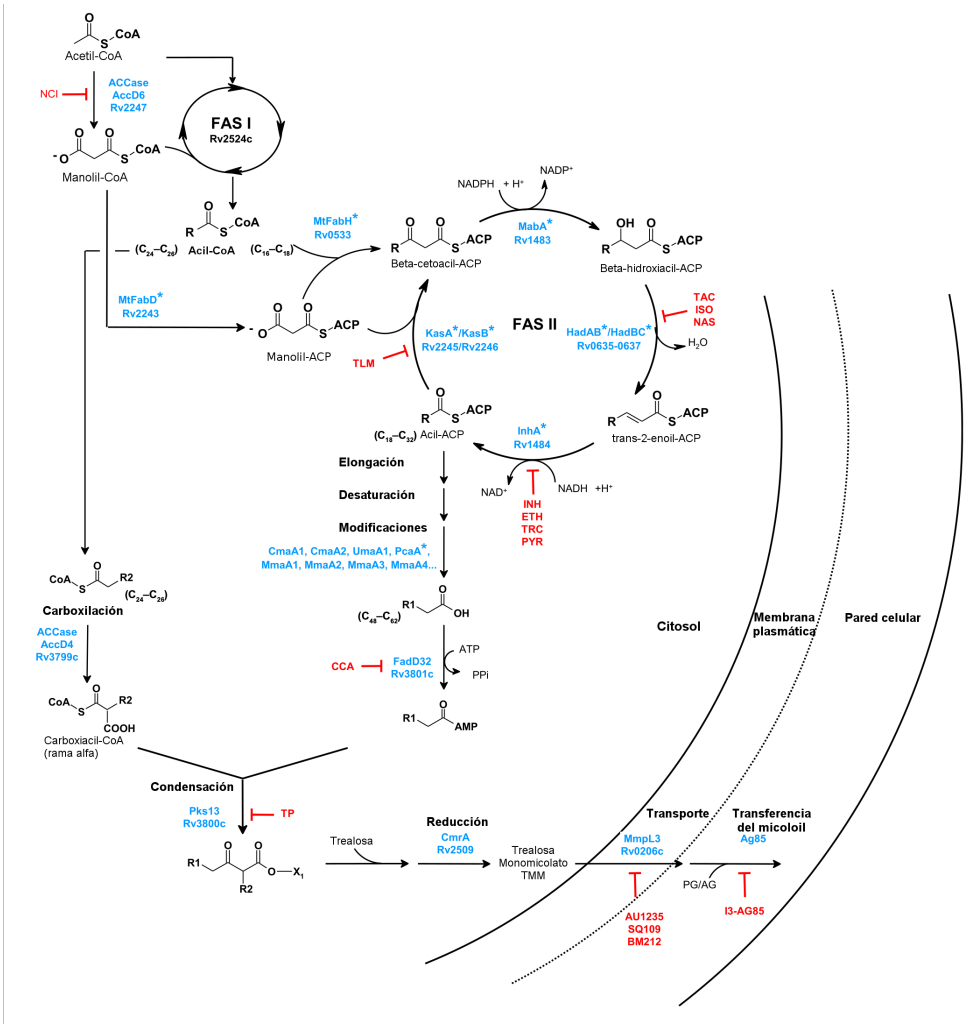


Figura 1.8: **Vía de síntesis del ácido micólico y su exportación a la pared**. En celeste están marcadas las proteínas involucradas en cada reacción. Las lineas rojas indican inhibición de ese paso por una de las drogas señaladas. Los asteriscos (*) indican que la proteína es fosforilada. Adaptado de (Marrakchi, Lanéelle, & Daffé, 2014)

.

1.4.3 *Mtb.*, la respuesta inmune y las ERON.

En la primera etapa de la infección con *Mtb.*, un macrófago alveolar lo fagocita. Dependiendo de distintas señales del ambiente la mycobacteria puede seguir en un estado replicativo o entrar en una fase de latencia, en el cual puede permanecer años. Si entra en el estadio replicativo produce necrosis y la enfermedad conocida como Tuberculosis. En la figura 1.9 se puede ver un resumen del proceso infectivo.

El Oxido Nítrico (NO) es el compuesto clave utilizado por el sistema inmunológico de los mamíferos para combatir infecciones. El mismo, es generado por los macrófagos, mediante la NO sintasa inducible (iNOS) (Voskuil et al., 2003) a través de la oxidación de la L-arginina dependiente de oxígeno y NADPH dando como resultado NO y citrunila (Marletta, 1993). El NO es liberado en el fagosoma donde debido a su alta permeabilidad penetra en la bacteria y ejerce su efecto. Si bien se sabe que el efecto del NO es principalmente bacteriostático, no se conoce en detalle su mecanismo de acción. Las propuestas del mecanismo de acción están asociadas a la química del NO donde se sabe que reacciona directamente con centros metálicos (Hemos, Cluster Fe-S y centros de Cu) y tioles. (Voskuil et al., 2003) También se ha propuesto que produce, mediante la reacción rápida con oxígeno y derivados de la cadena respiratoria, a quien inhibe parcialmente, diferentes especies reactivas de nitrógeno y oxígeno como ser el •OONO, el •NO2, y el •OH, todos agentes capaces de modificar postraduccionalmente las proteínas, particularmente en residuos Tirosina y/o Cisteina e inhibir su función. (Ischiropoulos, 2009; Miersch, & Mutus, 2005) El NO tiene un efecto bactericida, con lo cual, inhibir las proteínas que resultan blanco del NO resultaría en un efecto sinérgico potenciando el efecto del NO.

Muchos microorganismos, poseen mecanismos de defensa frente al NO y el estrés nitro/oxidativo asociado, que involucra la transcripción de genes específicos para eliminar los compuestos reactivos. (Gupta, & Chatterji, 2005) En particular, el bacilo de la tuberculosis es capaz de sobrevivir en el macrófago durante años, en lo que se conoce como fase de latencia, pudiendo desarrollar una infección posterior en el huésped. (Voskuil et al., 2003) En esta fase de latencia (también conocida como *Dormancy* en inglés) las drogas bacteriostáticas no son efectivas ya que el crecimiento es muy lento. El micro-entorno del bacilo en estas condiciones es conocido solo parcialmente suponiendo que el mismo está asociado a condiciones

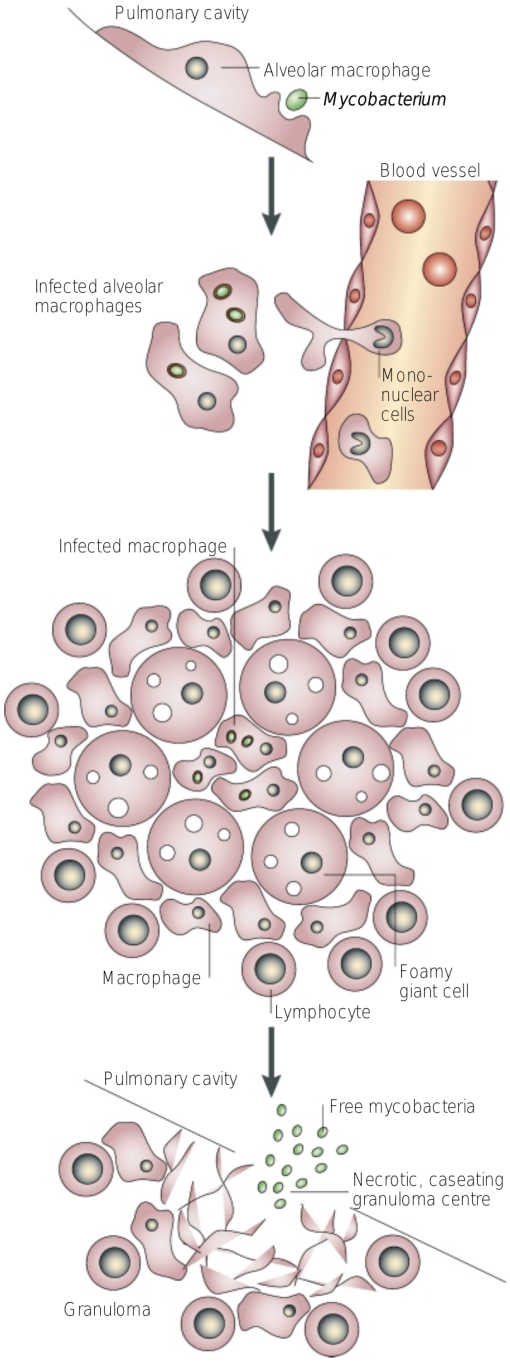


Figura 1.9: **Proceso de infección de *Mtb.*.** Una vez en los pulmones, cuyo ingreso se originó por aerosoles, el bacilo es fagocitado por un macrófago alveolar que induce una respuesta proinflamatoria. Este tipo de respuesta es la que define a la enfermedad, formando un tubérculo o granuloma, en donde los macrófagos infectados son rodeados por células espumosas gigantes (células derivadas de macrófagos fusionados, multinucleadas y con gotas lipídicas) y un manto exterior de linfocitos. En esta etapa la infección no se propaga. Ante alguna falla del sistema inmune, esta barrera de contención de rompe y se produce la infección. Tomado de (Russell, 2001)

de: hipoxia, acidez, escasez de nutrientes y la presencia de las especies reactivas mencionadas. La mycobacteria se vale de una serie de sensores para determinar el estado redox del medio en el que se encuentra. Los principales sistemas de sensado y transducción de señales son las

histidin-quinasas DosS y DosT (Bhat et al., 2012), se tratan de proteínas que tienen a su control proteínas de respuesta a estrés de ERON (regulon DosR), bajan la expresión de genes del metabolismo basal y provocan el ensanchamiento de la pared celular (Bhat et al., 2012). Estas proteínas se encuentran compuestas por dos dominios, un dominio sensor formado por una hemoproteína que reacciona con las ERON y los gases solubles permitiendo sensar al mismo tiempo el nivel redox del ambiente y la concentración de oxigeno (hipoxia); y otro dominio, el dominio transductor quinasa que fosforila un residuo de histidina en *cis*, activándose y fosforilando sus blancos moleculares, aproximadamente 50 genes que conforman el denominado regulón DosR. Los detonantes biológicos y moleculares del pasaje del estado de latencia a un estado infectivo son aún desconocidos por lo cual encontrar fármacos eficientes para la fase de latencia, y comprender la biología del bacilo en este estadio, es una de las necesidades imperiosas para combatir la Tuberculosis.

En resumen, *Mtb.* es un organismo intracelular, agente causante de la Tuberculosis. Durante su ciclo de vida intracelular experimenta una serie de condiciones de estrés que lo llevan a un estado conocido como de latencia o *dormancy*. Actualmente las drogas aprobadas para uso clínico no son efectivas en este estado, en el cual puede permanecer años, con lo cual es de suma importancia buscar fármacos específicos para este estadio infectivo. Los desarrollos de esta tesis si bien están centrados en las particularidades biológicas del mismo presentan una estrategia general que busca ir del genoma a la droga integrando una serie de metodologías bioinformáticas en pos de un objetivo común.

1.5 Objetivos e hipótesis de trabajo

Dada la importancia de las especies reactivas de oxígeno y nitrógeno (ERON) en la respuesta inmune del hospedador, nuestra hipótesis de trabajo busca atacar proteínas sensibles a este tipo de compuestos y que sean desde un punto de vista farmacológico relevantes para evitar el crecimiento y/o matar a *Mtb.*. Para poder poner a prueba dicha hipótesis planteo los siguientes objetivos:

- Desarrollar una herramienta que permita la elección de blancos moleculares basados en *bindability* y drogabilidad contextual, entendiendo contextual tanto desde un punto de vista metabólico como de estado infectivo en *Mycobacterium tuberculosis* utilizando in-

formación de expresión de microarreglos en hipoxia, estrés de ERON, falta de nutrientes, durante la infección de ratones y la sensibilidad de los blancos a ser modificados por NO.

- Elegir un blanco molecular basado en los criterios propuestos anteriormente y comprender su rol dentro de los metabolismos de *Mtb.* como así también su funcionamiento.

- Realizar búsquedas masivas virtuales de posibles compuestos tipo-droga para inhibir la actividad de la(s) enzima(s) y proponer familias de compuestos con posible actividad inhibitoria realizando los experimentos de inhibición/unión correspondientes.

A cada objetivo le corresponde un capítulo en donde se realiza una breve introducción al tema, se describen y discuten los métodos y resultados obtenidos.

Bibliografía

Abdallah, Abdallah M et al. (2007). "Type VII secretion—mycobacteria show the way". In: *Nature reviews microbiology* 5.11, pp. 883–891.

Agüero, Fernán et al. (2008). "Genomic-scale prioritization of drug targets: the TDR Targets database". In: *Nature Reviews Drug Discovery* 7.11, pp. 900–907.

Ananthan, Subramaniam et al. (2009). "High-throughput screening for inhibitors of Mycobacterium tuberculosis H37Rv". In: *Tuberculosis* 89.5, pp. 334–353.

Bank, Protein Data (1971). "Protein Data Bank". In: *Nature New Biol* 233, p. 223.

Barkan, Daniel et al. (2012). "Mycobacterium tuberculosis lacking all mycolic acid cyclopropanation is viable but highly attenuated and hyperinflammatory in mice". In: *Infection and immunity* 80.6, pp. 1958–1968.

Bateman, Alex et al. (2004). "The Pfam protein families database". In: *Nucleic acids research* 32.suppl 1, pp. D138–D141.

Besnard, Jérémy et al. (2012). "Automated design of ligands to polypharmacological profiles". In: *Nature* 492.7428, pp. 215–220.

Bhat, Shabir Ahmad et al. (2012). "The mechanism of redox sensing in Mycobacterium tuberculosis". In: *Free Radical Biology and Medicine* 53.8, pp. 1625–1641.

Biasini, Marco et al. (2014). "SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information". In: *Nucleic acids research*, gku340.

Bleicher, Konrad H et al. (2003). "Hit and lead generation: beyond high-throughput screening". In: *Nature Reviews Drug Discovery* 2.5, pp. 369–378.

Branchi, A et al. (1995). "Effects of bezafibrate and of 2 HMG-CoA reductase inhibitors on lipoprotein (a) level in hypercholesterolemic patients." In: *International journal of clinical pharmacology and therapeutics* 33.6, pp. 345–350.

Caminero, José A et al. (2010). “Best drug treatment for multidrug-resistant and extensively drug-resistant tuberculosis”. In: *The Lancet infectious diseases* 10.9, pp. 621–629.

Cheng, Alan C et al. (2007). “Structure-based maximal affinity model predicts small-molecule druggability”. In: *Nature biotechnology* 25.1, pp. 71–75.

Collins, LA, MN Torrero, & SG Franzblau (1998). “Green Fluorescent Protein Reporter Microplate Assay for High-Throughput Screening of Compounds againstMycobacterium tuberculosis”. In: *Antimicrobial agents and chemotherapy* 42.2, pp. 344–347.

Crick, Francis et al. (1961). “General nature of the genetic code for proteins”. In: *Nature* 192.4809, pp. 1227–32.

Davis, Fred P et al. (2007). “Host–pathogen protein interactions predicted by comparative modeling”. In: *Protein Science* 16.12, pp. 2585–2596.

Dayhoff, Margaret O (1965). “Computer aids to protein sequence determination”. In: *Journal of theoretical biology* 8.1, pp. 97–112.

Dayhoff, Margaret Oakley, & Robert S Ledley (1962). “Comprotein: a computer program to aid primary protein structure determination”. In: *Proceedings of the December 4-6, 1962, fall joint computer conference*. ACM, pp. 262–274.

Eswar, Narayanan et al. (2008). “Protein structure modeling with MODELLER”. In: *Structural Proteomics*. Springer, pp. 145–159.

Falchi, Federico et al. (2009). “3D QSAR Models Built on Structure-Based Alignments of Abl Tyrosine Kinase Inhibitors”. In: *ChemMedChem* 4.6, pp. 976–987.

Fujita, Toshio, & Corwin Hansch (1967). “Analysis of the structure-activity relationship of the sulfonamide drugs using substituent constants”. In: *Journal of medicinal chemistry* 10.6, pp. 991–1000.

Furin, J et al. (2001). “Occurrence of serious adverse effects in patients receiving community-based therapy for multidrug-resistant tuberculosis”. In: *The International Journal of Tuberculosis and Lung Disease* 5.7, pp. 648–655.

Gough, Julian et al. (2001). “Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure”. In: *Journal of molecular biology* 313.4, pp. 903–919.

Gupta, Surbhi, & Dipankar Chatterji (2005). “Stress responses in mycobacteria”. In: *IUBMB life* 57.3, pp. 149–159.

Haft, Daniel H, Jeremy D Selengut, & Owen White (2003). “The TIGRFAMs database of protein families”. In: *Nucleic acids research* 31.1, pp. 371–373.

Hagen, Joel B (2000). “The origins of bioinformatics”. In: *Nature Reviews Genetics* 1.3, pp. 231–236.

Halgren, Thomas A (2009). “Identifying and characterizing binding sites and assessing druggability”. In: *Journal of chemical information and modeling* 49.2, pp. 377–389.

Hopkins, Andrew L, & Colin R Groom (2002). “The druggable genome”. In: *Nature reviews Drug discovery* 1.9, pp. 727–730.

Huang, Dandan et al. (2012). “3D QSAR pharmacophore modeling for c-Met kinase inhibitors”. In: *Medicinal chemistry* 8.6, pp. 1117–1125.

Hulo, Nicolas et al. (2006). “The PROSITE database”. In: *Nucleic acids research* 34.suppl 1, pp. D227–D230.

Hutchinson, Lisa, & Rebecca Kirk (2011). “High drug attrition rates—where are we going wrong?” In: *Nature Reviews Clinical Oncology* 8.4, pp. 189–190.

Illergård, Kristoffer, David H Ardell, & Arne Elofsson (2009). “Structure is three to ten times more conserved than sequence—a study of structural response in protein cores”. In: *Proteins: Structure, Function, and Bioinformatics* 77.3, pp. 499–508.

Ischiropoulos, Harry (2009). “Protein tyrosine nitration—an update”. In: *Archives of Biochemistry and Biophysics* 484.2, pp. 117–121.

Kelley, Lawrence A et al. (2015). “The Phyre2 web portal for protein modeling, prediction and analysis”. In: *Nature protocols* 10.6, pp. 845–858.

Kinnings, Sarah L et al. (2010). “The Mycobacterium tuberculosis drugome and its polypharmacological implications”. In: *PLoS computational biology* 6.11, e1000976.

Kola, Ismail, & John Landis (2004). “Can the pharmaceutical industry reduce attrition rates?” In: *Nature reviews Drug discovery* 3.8, pp. 711–716.

Koul, Anil, Eric Arnoult, et al. (2011). “The challenge of new drug discovery for tuberculosis”. In: *Nature* 469.7331, pp. 483–490.

Koul, Anil, Luc Vranckx, et al. (2014). “Delayed bactericidal response of Mycobacterium tuberculosis to bedaquiline involves remodelling of bacterial metabolism”. In: *Nature communications* 5.

Lam, PY et al. (1994). “Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors”. In: *Science* 263.5145, pp. 380–384.

Lang, P Therese et al. (2009). “DOCK 6: Combining techniques to model RNA–small molecule complexes”. In: *Rna* 15.6, pp. 1219–1230.

Letunic, Ivica, Tobias Doerks, & Peer Bork (2012). “SMART 7: recent updates to the protein domain annotation resource”. In: *Nucleic acids research* 40.D1, pp. D302–D305.

Lew, Jocelyne M et al. (2011). “TubercuList–10 years after”. In: *Tuberculosis* 91.1, pp. 1–7.

Marletta, Michael A (1993). *Nitric oxide synthase structure and mechanism*. ASBMB.

Marrakchi, Hedia, Marie-Antoinette Lanéelle, & Mamadou Daffé (2014). “Mycolic acids: structures, biosynthesis, and beyond”. In: *Chemistry & biology* 21.1, pp. 67–85.

Miersch, Shane, & Bulent Mutus (2005). “Protein S-nitrosation: biochemistry and characterization of protein thiol–NO interactions as cellular signals”. In: *Clinical biochemistry* 38.9, pp. 777–791.

Morris, Garrett M et al. (2009). “AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility”. In: *Journal of computational chemistry* 30.16, pp. 2785–2791.

Mullard, Asher (2015). “The phenotypic screening pendulum swings”. In: *Nature Reviews Drug Discovery* 14.12, pp. 807–809.

— (2016). “2015 FDA drug approvals”. In: *Nature Reviews Drug Discovery* 15.2, pp. 73–76.

Organization, World Health et al. (2010). “Guidelines for treatment of tuberculosis”. In: *World Health Organization, Geneva, Switzerland*.

Pausch, Mark H (1997). “G-protein-coupled receptors in Saccharomyces cerevisiae: high-throughput screening assays for drug discovery”. In: *Trends in biotechnology* 15.12, pp. 487–494.

Pieper, Ursula et al. (2011). “ModBase, a database of annotated comparative protein structure models, and associated resources”. In: *Nucleic acids research* 39.suppl 1, pp. D465–D474.

Reddy, TBK et al. (2009). “TB database: an integrated platform for tuberculosis research”. In: *Nucleic acids research* 37.suppl 1, pp. D499–D508.

Roses, Allen D (2008). "Pharmacogenetics in drug discovery and development: a translational perspective". In: *Nature Reviews Drug Discovery* 7.10, pp. 807–817.

Ruiz-Carmona, Sergio et al. (2014). "rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids". In: *PLoS Comput Biol* 10.4, e1003571.

Russell, David G (2001). "Mycobacterium tuberculosis: here today, and here tomorrow". In: *Nature Reviews Molecular Cell Biology* 2.8, pp. 569–586.

Russell, David G, Clifton E Barry, & JoAnne L Flynn (2010). "Tuberculosis: what we don't know can, and does, hurt us". In: *Science* 328.5980, pp. 852–856.

Sanger, Frederick, Steven Nicklen, & Alan R Coulson (1977). "DNA sequencing with chain-terminating inhibitors". In: *Proceedings of the National Academy of Sciences* 74.12, pp. 5463–5467.

Schilling, Christophe H et al. (1999). "Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era". In: *Biotechnology progress* 15.3, pp. 296–303.

Schmidtke, Peter, & Xavier Barril (2010). "Understanding and predicting druggability. A high-throughput method for detection of drug binding sites". In: *Journal of medicinal chemistry* 53.15, pp. 5858–5867.

Schultz, Jörg et al. (1998). "SMART, a simple modular architecture research tool: identification of signaling domains". In: *Proceedings of the National Academy of Sciences* 95.11, pp. 5857–5864.

Sprous, DG et al. (2006). "Kinase inhibitor recognition by use of a multivariable QSAR model". In: *Journal of Molecular Graphics and Modelling* 24.4, pp. 278–295.

Sundberg, Steven A (2000). "High-throughput and ultra-high-throughput screening: solution- and cell-based approaches". In: *Current opinion in biotechnology* 11.1, pp. 47–53.

Thomas, Paul D et al. (2003). "PANTHER: a library of protein families and subfamilies indexed by function". In: *Genome research* 13.9, pp. 2129–2141.

Voskuil, Martin I et al. (2003). "Inhibition of respiration by nitric oxide induces a Mycobacterium tuberculosis dormancy program". In: *The Journal of experimental medicine* 198.5, pp. 705–713.

WHO (2014). "Global tuberculosis report 2014". In: p. 171. URL: `http://apps.who.int/iris/handle/10665/91355`.

Wlodawer, Alexander, & Jiri Vondrasek (1998). “INHIBITORS OF HIV-1 PROTEASE: A Major Success of Structure-Assisted Drug Design 1”. In: *Annual review of biophysics and biomolecular structure* 27.1, pp. 249–284.

Wu, Cathy H et al. (2004). “PIRSF: family classification system at the Protein Information Resource”. In: *Nucleic acids research* 32.suppl 1, pp. D112–D114.

Xie, Lei, & Philip E Bourne (2007). “A robust and efficient algorithm for the shape description of protein structures and its application in predicting ligand binding sites”. In: *BMC bioinformatics* 8.Suppl 4, S9.

Yeats, Corin et al. (2006). “Gene3D: modelling protein structure, function and evolution”. In: *Nucleic acids research* 34.suppl 1, pp. D281–D284.

Capítulo 2

Métodos computacionales

En esta capítulo se hará una breve introducción teórica a los métodos de Química Computacional utilizados en esta tesis. En cada capítulo específico se comentarán en detalle los protocolos utilizados para realizar las simulaciones. Se realizaron tres tipos de simulaciones: i)*Docking*, ii) Dinámica molecular clásica y iii) Dinámica molecular híbrida (QM-MM). La dinámica molecular clásica se utilizó como base para los cálculos híbridos y como herramienta de re puntuación de las poses de *docking* utilizando la descomposición de la energía mediante MM-PBSA. Por esta razón comenzaremos comentando esta familia de técnicas. Luego comentaremos los métodos cuánticos, en particular el hamiltoniano semi-empírico DFTB (Density Fuctional Tight Binding) para después introducir las metodologías híbridas (QM-MM). Por último, daremos una introducción a las técnicas de *docking* poniendo especial énfasis en las específicamente implementadas por el programa rDock.

2.1 Dinámica molecular clásica

Muchos sistemas de interés en química y biología requieren del estudio molecular de sistemas de gran cantidad de átomos, como por ejemplo las proteínas. Para este tipo de sistemas resulta, aún con el poder computacional existente hoy en día, prácticamente imposible su tratamiento completo con mecánica cuántica. Es por esto que en gran cantidad de problemas en los que no se requiere un detalle de la distribución electrónica, se utilizan métodos basados en la mecánica clásica, lo que se denomina mecánica molecular (MM). En estos métodos se ignora el movimiento de los electrones, y se calcula la energía exclusivamente en función de las posi-

ciones de los núcleos. Sin embargo, la mecánica cuántica se utiliza para el desarrollo de los distintos parámetros requeridos para el cálculo, que se explicarán en la próxima sección.

2.1.1 Campos de fuerza clásicos

En los métodos de mecánica molecular, la energía potencial viene dada por lo que se denomina campo de fuerza, que no es más que una expresión de la energía potencial dependiente de las coordenadas de los núcleos y una serie de parámetros. La expresión del campo de fuerza AMBER 99SB (Hornak et al., 2006) - el cual fue utilizado para los cálculos realizados en este trabajo - contiene dos clases de contribuciones, las contribuciones de unión y las de no unión. Las de unión, se calculan para átomos formando un enlace o como máximo a 2 átomos de distancia, e incluyen penalidades para las desviaciones respecto de su valor de equilibrio de distancias de enlace, ángulos y ángulos diedros. Mientras que las dos primeras se representan a través de un potencial armónico centrado en la distancia de equilibrio, los ángulos diedros o torsiones están descriptos con una función periódica. Las contribuciones de no unión, se computan entre átomos que están en diferentes moléculas o a más de 4 enlaces de distancia, y se dividen en interacciones electrostáticas basadas en la ley de Coulomb, por un lado, y un potencial de Lennard-Jones, que describe en forma aproximada las interacciones dispersivas y la repulsión originada por el principio de exclusión de Pauli a distancias muy cortas. La expresión de la energía potencial resulta en el caso del campo de fuerzas AMBER utilizado en esta tesis en:

$$E(r^N) = \sum_{enlaces} \frac{k_{b,i}}{2}(r_i - r_{i0}) + \sum_{angulos} \frac{k_{\theta,i}}{2}(\theta_i - \theta_{i0}) + \sum_{torsiones} \frac{V_n}{2}\cos(n\varpi - \gamma) + \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left\{\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\right\} \tag{2.1}$$

Comprendamos en detalle cada una de las contribuciones al campo de fuerza clásico. El primer término modela la contribución de estiramiento de los enlaces, el cual se describe a través de un potencial armónico. En éste, $k_{b,i}$ corresponde a la constante de fuerza asociada a la unión i y r_{io} a la distancia de equilibrio de la misma unión. Si bien una mejor descripción de un enlace químico se obtendría con un potencial como por ejemplo el sugerido por Morse, este

contiene un parámetro adicional y no es tan comúnmente utilizado en cálculos MM debido a su costo adicional. Por otra parte, el potencial armónico resulta una buena y sencilla aproximación al potencial real en la zona cercana al mínimo, donde no hay ruptura de enlaces. El segundo término corresponde a las contribuciones dadas por las flexiones angulares. Estas también se representan con un potencial armónico de constante $k_{\theta,i}$ y valor de equilibro θ_0. Normalmente, la energía requerida para modificar un ángulo es menor a la requerida para variar la distancia entre dos átomos, por lo que los valores de $k_{\theta,i}$ suelen ser menores que los de las constantes $k_{b,i}$. El término asociado a las torsiones o ángulos diedros presenta tres parámetros: V_n corresponde a la barrera energética asociada a la rotación entre dos mínimos sucesivos, n representa la multiplicidad del potencial, es decir indica el número de mínimos encontrados entre 0°y 360°, y por último γ, que corresponde a la fase de la función sinusoidal.

El último término de la ecuación 2.1 contiene las contribuciones de no unión, y depende de la distancia entre pares de átomos, r_{ij}. La primera parte, corresponde al potencial de Van der Waals, y contiene los parámetros ε_{ij}, asociado a la profundidad del pozo de energía, y σ_{ij}, correspondiente a la distancia entre los átomos correspondiente al mínimo de energía. La parte electrostática contiene como parámetros las cargas sobre los átomos q_i y q_j. En el campo de fuerzas de AMBER, la distribución de cargas es representada a través de cargas puntuales ubicadas en el centro de las coordenadas de cada átomo. Estas cargas se obtienen de manera de reproducir el potencial electrostático de la molécula, a partir de cálculos cuánticos. Los valores de las cargas parciales se realiza a través del computo de una grilla de potencial electrostático determinada utilizando un hamiltoniano cuántico (sea este Hartree-Fock/6-31G* o AM1-bcc) ajustando los valores de las cargas puntuales en dos pasos, obteniendo primero los valores de los heteroátomos y, restringiendo el valor de los heteroátomos al obtenido en el paso anterior, de los hidrógenos. (Bayly et al., 1993; Cieplak et al., 1995)

Para disminuir el costo computacional asociado a los términos de no unión, se utiliza un radio de corte, tal que para los átomos que se encuentran a una distancia mayor que este radio de corte estas interacciones no se computan. En el caso de las interacciones de Lennard-Jones, al ser de corto alcance, esto no genera un problema siempre que el radio de corte sea lo suficientemente grande. En el caso de las interacciones coulómbicas, de mayor alcance, en los casos en que se utilizan condiciones periódicas de contorno en la simulación se utiliza una metodología

denominada sumas de Ewald, un método originalmente desarrollado para el estudio de cristales iónicos, en su implementación denominada Particle Mesh Ewald (PME)(Darden, York, & Pedersen, 1993; Salomon-Ferrer et al., 2013). Esta metodología se utiliza para calcular las interacciones electrostáticas totales entre los átomos presentes en cada celda unidad, de forma muy eficiente. La energía electrostática se calcula en dos partes: la primera parte, en la cual a las interacciones electrostáticas de corto alcance se calculan en el espacio real mientras que la segunda parte (en el espacio reciproco) se calcula utilizando la transformada de Fourier. Este procedimiento tiene como ventaja una convergencia mucha más rápida que utilizando cálculos de interacciones directas. De esta manera se logra tener en cuenta las interacciones electrostáticas de largo alcance más allá del radio de corte de las interacciones de no unión, en forma altamente eficiente.

Finalmente, puede observarse que la mera expresión de la energía potencial no es suficiente para definir un campo de fuerza. El campo de fuerzas es constituido además por la gran cantidad de parámetros asociados a cada uno de los sumandos en la energía potencial. Para el caso de simulación de proteínas, el campo de fuerza de AMBER contiene los parámetros asociados a los aminoácidos naturales presentes en la mayoría de las proteínas, en sus distintos estados de oxidación o protonación, cuando corresponda. En este contexto, puede hablarse de la transferibilidad del campo de fuerza. Es decir, el hecho de que el mismo conjunto de parámetros puede utilizarse para moléculas asociadas. Por ejemplo, los parámetros para un residuo de triptofano serán los mismos en todas las proteínas estudiadas, y no será necesaria su parametrización en cada proteína que se desee estudiar. En el campo de fuerza de AMBER, los átomos dentro de cada residuo reciben un nombre de átomo y un tipo de átomo. Los tipos de átomo permiten que no sea necesario incluir parámetros para cada átomo incluido en cada uno de los aminoácidos. Distintas especies pueden poseer el mismo tipo de átomo, y los parámetros asociados a las uniones, los ángulos y los diedros (los denominados parámetros de unión) se dan por tipo de átomo, en lugar de para cada átomo de cada residuo en particular. Si bien en el campo de fuerza se incluyen parámetros específicos para todos los aminoácidos naturales y otras moléculas comúnmente encontradas en los sistemas biológicos, para incluir modificaciones post-traduccionales como la fosforilación, será necesario obtener los parámetros necesarios para realizar la simulación. El concepto de transferibilidad del campo de fuerza se utiliza

a la hora de simular compuestos tipo droga y no tener que obtener de cero los parámetros de enlace (distancias de enlace, ángulos, diedros y ángulos impropios). En este trabajo utilizamos la estrategia desarrollada en el *Generalized AMBER Force Field* (GAFF). (Wang et al., 2004) sólo siendo necesario derivar las cargas parciales mediante cálculos cuánticos y el algoritmo RESP (*Restraint Electrostatic Potential*).

Por último, cabe realizar un comentario especial referido a la forma de representación de las moléculas de agua en el campo de fuerza. La manera de describir las moléculas de agua representa un factor importante en el cálculo, dada la gran cantidad de moléculas de agua normalmente presentes en el sistema. En los cálculos realizados en este trabajo, se utilizó un modelo de agua simple denominado TIP3P. En los modelos de agua simples, cada molécula de agua se mantiene en una geometría rígida, y la interacción entre moléculas de agua se describe a través de interacciones Coulómbicas y de Lennard-Jones. En el modelo TIP3P se ubican tres cargas puntuales, una sobre cada átomo de la molécula. La carga negativa ubicada en el átomo de oxígeno es compensada por las cargas positivas en los átomos de hidrógeno. La distancia entre el oxígeno y los átomos de hidrógeno, así como el ángulo H-O-H se mantienen fijos (mediante el uso del algoritmo de SHAKE (SETTLE, 1992). Otros modelos de agua simples usualmente utilizados son por ejemplo el modelo SPC/2 o su actualización SPC/2, y el modelo TIP4P, que utiliza cuatro puntos en vez de tres para describir la distribución de cargas. Los valores de los parámetros asociados a los modelos de agua simples se desarrollan de forma de reproducir las propiedades del agua medidas experimentalmente como la densidad, la función de distribución radial, y otras propiedades fisicoquímicas. Cómo se tratan de modelos sencillos, existen gran cantidad de propiedades que no pueden ser descriptas. Existen modelos más complejos, como así también más costosos, que incluyen efectos de polarización, lo que resulta de importancia en sistemas donde se espera que el solvente experimente un efecto de polarización significativo por parte del resto del sistema.

2.1.2 Dinámica molecular

Hasta ahora hemos explicado como se construye el potencial clásico pero no como se calcula el movimiento ni que otros elementos son necesarios para realizar las simulaciones en fase condensada. Para realizar los movimientos en cada paso de integración (en nuestro caso 2

femtosegundos para dinámicas clásicas y 1 -0.5 fs. para los cálculos QM-MM) se utiliza un algoritmo conocida como Velocity Verlet (Grubmüller et al., 1991) que fue desarrollado para resolver las ecuaciones de movimiento de Newton de forma eficiente permitiendo calcular las velocidades (necesarias para determinar la energía cinética y, por lo tanto la temperatura) y posición de los átomos en el mismo paso de integración. Otro elemento relevante para realizar las simulaciones en un ensamble Isotérmico-Isobárico (NPT), relevante para sistemas de interés biológico es tener tanto un termostato como un barostato para mantener estas propiedades oscilando en torno a un valor (300K y 1 bar). El termostato utilizado para las simulaciones de este trabajo fue el termostato de Berendsen (Berendsen et al., 1984). Dicho termostato utiliza un decaimiento exponencial de las fluctuaciones de la energía cinética (a una temperatura T_0 con una constante de tiempo (τ). (Ver ecuación 2.2)

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau} \tag{2.2}$$

Lo mismo sucede con la presión, dado que los sistemas biológicos se encuentran a presión constante es necesario conservar esta propiedad oscilando en un valor. Para ello se utilizó la misma estrategia realizando las simulaciones en presencia del barostato de Berendsen.

2.2 La mecánica cuántica

No fue posible hasta el desarrollo de la mecánica cuántica la obtención de parámetros precisos para poder modelar adecuadamente el comportamiento de átomos y moléculas. El trabajo de Max Planck al introducir por primera vez la noción de que la energía (o el espectro de emisión de un sólido) no tiene que ser necesariamente continua como sostenía la física clásica newtoniana sentó las bases para las posteriores formulaciones de De Broglie, Heissenberg y Schröedinger. A partir de allí, sería la ecuación de Schröedinger la que describiría el comportamiento de átomos y moléculas y daría origen a la mecánica cuántica:

$$-\hbar\frac{\partial}{\partial t}\Psi = \hat{H}\Psi \tag{2.3}$$

En el desarrollo de este trabajo de tesis fue necesario utilizar resultados derivados de la resolución aproximada de la Ecuación de Schröedinger utilizando la teoría del funcional de la

densidad (DFT, por sus siglas en inglés) y su aproximación, DFTB (Density Functional Tight Binding) como así también Hartree-Fock (HF).

La teoría del funcional de la densidad permite, mediante el uso de funcionales (funciones de funciones), calcular las propiedades electrónicas de un sistema, que dependen espacialmente de la densidad electrónica. La energía total de un sistema en DFT se descompone en varias componentes:

$$E[n(r)] = T_s + E_{ext} + E_H + E_{ex} + E_{II} \tag{2.4}$$

donde T_s representa la energía cinética de no interacción, E_{ext} es la energía externa de interacción (entre los electrones y los núcleos), E_{II} es la energía de interacción entre los núcleos, E_H es la energía de Hartree y E_{ex} es la energía de intercambio y correlación donde quedan ocultos todos los efectos de muchos cuerpos difíciles de computar. La dificultad de resolver la estructura electrónica radica en el término de intercambio y correlación.

2.2.1 SCC-DFTB

El Self-Consistent Charge, Density Functional Tight-Binding (SCC-DFTB) es una aproximación metodológica basada en la Teoría del Funcional de la Densidad (DFT), se trata de un método comúnmente enmarcado de la familia de los hamiltonianos semiempíricos, que tienen una porción parametrizada sobre datos experimentales (que acelera mucho el cómputo) y otra que se calcula *ab initio* a partir de aproximaciones. En la aproximación SCC-DFTB la densidad electrónica de la teoría DFT es sustituida por una densidad de referencia más las fluctuaciones de la misma $\rho = \rho_0(r) + \delta_\rho(r)$. La energía total de DFT es expandida hasta un segundo orden para las fluctuaciones de la densidad de carga. Luego de una serie de aproximaciones la energía puede ser escrita de la siguiente forma (Marcus Elstner et al., 1998):

$$E^{SCC-DFTB} = \sum_{i}^{OCC} \left\langle \psi_i \left| \hat{H}_0 \right| \psi_i \right\rangle + E_{rep} + \frac{1}{2} \sum_{\alpha\beta} \Delta q_\alpha \Delta q_\beta \gamma_{\alpha\beta} \tag{2.5}$$

Los primeros dos términos de la ecuación 2.5 son computados en la densidad de referencia ρ_0 y forman la aproximación original de DFTB. (Porezag et al., 1995) El primer término es el ”término de la estructura de banda” (esta terminología deriva de la ciencia de materiales de

donde la aproximación DFTB se origina), y se refiere al elemento *leading* del hamiltoniano, H0. Los orbitales moleculares de Kohn-Sham ϕ_i son expandidos en unas bases mínimas confinadas a orbitales atómicos del tipo Slater por ejemplo: $\phi_i = \sum_V C_{VI} \emptyset_V$ como fue descripto por (Eschrig, & Bergert, 1978) que es determinada resolviendo el problema atómico de Kohn-Sham en presencia de un potencial de confinamiento (Porezag et al., 1995). Estos orbitales atómicos (ω_v) son luego usados para calcular los elementos de matriz del Hamiltoniano como:

$$H^0_{\mu\nu} = \begin{cases} \varepsilon_\mu^{free\ atom},\ if\ \ \phi_\mu = \phi_\nu\ ; \\ \left\langle \phi_\mu \left| \hat{H}(\rho^0_\alpha + \rho^0_\beta) \right| \phi_\nu \right\rangle,\ \ \phi_\mu \in \alpha\ y\ \phi_\nu \in \beta\ ; \\ 0\ , en\ todos\ los\ otros\ casos \end{cases} \tag{2.6}$$

En la ecuación 2.6 $\varepsilon_\mu^{free\ atom}$ es el autovalor de Kohn-Sham para el orbital ρ_μ en el átomo sin confinamiento, y los tres términos del centro han sido descartados. El Hamiltoniano y el solapamiento de los elementos de la matriz son entonces pre-calculados en esta aproximación de dos centros para las distancias interatómicas en una escala relevante y tabuladas (Porezag et al., 1995; Zhechkov et al., 2005) y elementos a distancias arbitrarias se obtienen al interpolar sobre los valores tabulados. El segundo término es el de interacción repulsiva de a pares, aproximada como la suma del potencial de dos cuerpos:

$$E_{rep} = \sum_{\alpha\beta} U_{\alpha\beta} \tag{2.7}$$

En la práctica, este termino es ajustado (utilizando una función tipo *spline*) a la diferencia de la energía total de un cálculo completo de DFT con una base mínima y la parte electrónica de un cálculo de DFTB (es decir, todos los términos salvo la E_{rep}) con respecto al largo del enlace de interés en una molécula de interés relevante. (M Elstner, 2006) El último término de la ecuación 2.5 define la aproximación de SCC de segundo orden. Introduce un término correctivo para las interacciones Coulombicas de larga distancia a la energía total, debido a las fluctuaciones en las cargas $\delta q_\alpha = q_\alpha - q^0_\alpha$ centradas en los átomos. El $\gamma_{\alpha\beta}$ es un función analítica que brinda un comportamiento correcto en los casos límite: para $R_\alpha = R_\beta, \gamma_{\alpha\beta}$ brinda la contribución por la autointeracción de α, que evalúa al parámetro de Hubbard del átomo $\alpha(U_\alpha)$ y en el límite de las largas distancias la interacción Coulombica entre dos esferas con una distribución de carga

centrada en R_α y R_{beta}.

Aplicando el principio variacional, que implica hallar mediante densidades de prueba la densidad electrónica de más baja energía y es el mismo que el utilizado en HF o DFT, la ecuación de Kohn-Sham final puede ser escrita como el siguiente sistema de ecuaciones algebraicas:

$$H_{\mu\nu} = \left\langle \phi_\mu \left| \hat{H}_0 \right| \phi_\nu \right\rangle + \frac{1}{2} S_{\mu\nu} \sum_{\zeta}^{N} (\gamma_{\alpha\zeta} + \gamma_{\beta\zeta}) \Delta q_\zeta \, \hat{H}^0_{\mu\nu} + \hat{H}^1_{\mu\nu}; \ \forall \mu \in \alpha, \nu \in \beta \tag{2.8}$$

Donde:

$$H_{\mu\nu} = \left\langle \phi_\mu \left| \hat{H}_0 \right| \phi_\nu \right\rangle ; S_{\mu\nu} = \left\langle \phi_\mu | \phi_\nu \right\rangle ; \forall \mu \in \alpha, \nu \in \beta \tag{2.9}$$

Por lo tanto, la corrección de segundo orden debido a la fluctuación de las cargas es representada por los elementos no diagonales $H^1_{\mu\nu}$, que dependen de la carga atómica. Estas cargas son calculadas por un análisis de Mulliken, y dependen del coeficiente $c_{\mu i}$ de los orbitales de Kohn-Sham. Entonces, el proceso debe ser iterado hasta que se adquiera una autoconsistencia. Como no es necesario calcular ninguna integral de solapamiento porque se encuentran pre-calculadas, el costo computacional esta dominado por la solución del problema de los autovalores de la ecuación 2.8.

Finalmente, una simple expresión analítica para las fuerzas interatómicas puede ser derivada tomando la derivada de la energía de SCC-DFTB respecto a las coordenadas nucleares para brindar:

$$F_\alpha = -\sum_{i}^{occ} n_i \sum_{\mu\nu} c_{\mu i} c_{\nu i} \left(\frac{\partial H^0_{\mu\nu}}{\partial R_\alpha} - (\epsilon_i - \frac{H^1_{\mu\nu}}{S_{\mu\nu}}) \frac{\partial S_{\mu\nu}}{\partial R_\alpha} \right) - \Delta q_\alpha \sum_{\zeta}^{QM} \frac{\partial \gamma_{\alpha\zeta}}{\partial R_\alpha} \Delta q_\zeta - \frac{\partial E_{rep}}{\partial R_\alpha} \tag{2.10}$$

Las derivadas para el Hamiltoniano y el solapamiento son calculadas si son necesarias a partir de los valores tabulados a través de diferencias finitas, mientras que fórmulas analíticas simples pueden obtenerse para los términos restantes.

2.3 Dinámica molecular híbrida (QM/MM)

Los métodos basados en la mecánica cuántica presentan información de la distribución electrónica del sistema y son imprescindibles cuando se desea estudiar procesos reactivos, pero traen apare-

jado un gran costo computacional. Por otro lado, los métodos llamados de mecánica molecular clásica (basados en campos de fuerza) presentan un costo computacional mucho menor, y permiten estudiar fenómenos estructurales asociados a moléculas de gran cantidad de átomos, procesos con movimientos característicos en la escala de los nanosegundos a diferencia de las técnicas cuánticas que se restringen a fenómenos que ocurren en la escala de los picosegundos. Se puede pensar que pueden combinarse ambas metodologías para el estudio de un sistema en el que sea necesaria una descripción de la densidad electrónica de una cierta región, que se debe describir con mecánica cuántica, y el resto del sistema puede tratarse con mecánica clásica. Un ejemplo de aplicación de este tipo de metodologías es el estudio de reacciones enzimáticas. (Defelipe et al., 2015; Crespo, Scherlis, et al., 2003; Crespo, Martí, Kalko, et al., 2005; Turjanski, Hummer, & Gutkind, 2009; Dumas et al., 2014; Arcon et al., 2015).

En ellas, el sitio reactivo debe tratarse necesariamente con mecánica cuántica, pero el resto de la enzima puede tratarse con mecánica clásica. Esto constituye un avance significativo respecto de la utilización de sistemas modelo para tratar este tipo de problemáticas, en los que solo se incluía el sitio activo y algunos grupos pequeños relevantes. En las metodologías híbridas efecto tanto de la polarización electrostática como estérico del entorno proteico es considerado explícitamente en la mayoría de los casos crucial en el desarrollo del fenómeno reactivo. Para aplicar una metodología QM-MM, en primer lugar es necesario definir el sector del sistema que se tratará con mecánica cuántica, que llamaremos subsistema QM. El resto del sistema subsistema MM se tratará con un campo de fuerza clásico. Para la elección, debe tenerse en cuenta que cuanto mayor sea el subsistema QM, mayor calidad tendrá el cálculo, pero también su costo se incrementará significativamente. En el caso de que la simulación a realizar consista en el estudio de una molécula inmersa en un solvente, resulta natural elegir como subsistema QM a la molécula a estudiar y tratar al solvente clásicamente. En el caso del estudio de una reacción enzimática, la elección es más compleja. Resulta imprescindible incluir en el subsistema QM todos los átomos que incluyan enlaces que puedan romperse o formarse durante el proceso de interés. Sin embargo, para obtener una buena representación del problema de estudio, normalmente se requiere incluir una mayor cantidad de átomos que simplemente los involucrados en la reacción propiamente dicha.

Dentro de los métodos QM-MM, pueden distinguirse dos clases de metodologías: los esque-

mas aditivos y los esquemas sustractivos. La metodología utilizada en este trabajo corresponde a un esquema aditivo, por lo que nos concentraremos en este tipo de esquemas. En los métodos QM-MM aditivos, el Hamiltoniano consiste en la suma de la contribución del subsistema QM (H_{QM}), el subsistema MM (H_{MM}) y un término de acoplamiento QM-MM (H_{QM-MM}): -

$$H = H_{QM} + H_{MM} + H_{QM-MM} \tag{2.11}$$

Para el cálculo de la energía E_{QM} asociada al H_{QM}, debe seleccionarse un método cuántico. Los cálculos presentados en este trabajo están realizados a nivel de DFTB (Density Functional based Tight Binding), un Hamiltoniano semiempírico. En el Hamiltoniano cuántico no solo se tienen en cuenta las cargas de los núcleos cuánticos, sino que se realiza el cálculo teniendo en cuenta además el potencial electrostático generado por las cargas clásicas. La energía E_{MM} se calcula a través del uso de un campo de fuerza. En el caso de este trabajo, el campo de fuerza utilizado fue el campo de fuerza AMBER99SB (Hornak et al., 2006), explicado en la sección anterior. El término clave en el cálculo QM-MM es el término de interacción H_{QM-MM}. La forma en que se define éste, da cuenta del método QM-MM particular. En forma general, podemos decir que incluye las interacciones electrostáticas, de van der Waals y de unión de los átomos en la frontera, entre los átomos QM y los MM. En el método utilizado en los cálculos realizados en este trabajo, el término de acoplamiento QM-MM consiste en el descripto en la expresión 2.12 para un sistema de A átomos MM y B átomos QM:

$$E_{QM-MM} = \sum_{i=1}^{A} q_i \int \frac{\rho(r)}{|r-\tau_i|} dr + \sum_{i=1}^{A}\sum_{j=1}^{B} \frac{q_i Z_j}{|R_j-\tau_i|} + \sum_{i=1}^{A}\sum_{j=1}^{B} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{\tau_i - R_j}\right)^{12} - \left(\frac{\sigma_{ij}}{\tau_i - R_j}\right)^{6}\right] \tag{2.12}$$

En la ecuación 2.12 τ_i corresponde a las posiciones de los núcleos MM, R_j a las coordenadas de los núcleos QM. El primer sumando da cuenta de la interacción electrostática entre la densidad electrónica del subsistema cuántico y las cargas sobre los átomos MM (q_i). El segundo término describe la interacción electrostática entre los núcleos QM, de carga Z_j, y los átomos MM. El tercer término describe las interacciones de van der Waals, en la misma forma que lo realiza en campo de fuerza clásico, descripto en la sección anterior. Este último término implica que sea requerido obtener los parámetros asociados al potencial de Lennard-Jones ε y

σ para los átomos del subsistema QM, los que normalmente son obtenidos del campo de fuerza utilizado para representar el subsistema MM.

2.4 Métodos de estimación de energía libre

Dado que para estimar la energía libre de un proceso se requiere de múltiples observaciones de dicho fenómeno (la ruptura y formación de enlaces, un cambio conformacional) a lo largo de una coordenada de reacción que describe el proceso (una distancia, un ángulo, una combinación lineal de distancias, etc). Si la barrera a atravesar es grande (mayor a k_B, será difícil observar siquiera una vez el proceso. Por esta razón para estudiar este tipo de fenómenos se recurren a métodos de muestreo sesgado.

2.4.1 Inigualdad de Jarzynski

El método de Dinámica Molecular Dirigida (MSMD, Multiple Steered Molecular Dynamics, por sus siglas en inglés) está basado en la igualdad demostrada por Jarzynski en 1997. (Jarzynski, 1997). De la termodinámica básica, sabemos qu en un sistema cerrado conectado a un reservorio térmico, el trabajo que se realice o sea entregado por el sistema para realizar un proceso que lo lleva de un estado inicial A a un estado final B, es mayor o igual al cambio en energía libre del sistema, cumpliéndose la igualdad solamente en el caso de que el trabajo sea realizado en forma reversible. Jarzynski demostró que la energía libre del proceso A B, puede obtenerse del promedio exponencial realizado sobre el ensamble formado por infinitas determinaciones de trabajo irreversible asociadas a caminos que conectan los estados A y B. La ecuación 2.13 muestra esta relación:

$$< e^{-\beta W} >_{A \to B} \geq e^{\frac{-\Delta G}{k_B T}} \tag{2.13}$$

donde W, el trabajo para llevar al sistema del estado A al B, es tomado de medidas realizadas desde las condiciones iniciales para el sistema en el estado A generadas en el ensamble canónico a una temperatura T.

La aplicación de esta ecuación al cálculo de perfiles de energía libre se realiza, al igual que con otros métodos como *Umbrella Sampling*, a través del agregado de un término armónico a

la energía potencial.(Martí et al., 2008). Pero, al contrario de este método, el mínimo de este potencial se mueve a lo largo de la coordenada de reacción durante la SMD como:

$$E(r) = E(r) + k[\zeta - (\zeta_0 + \nu\Delta t)]^2 \tag{2.14}$$

donde ν corresponde a la velocidad a la que se mueve la coordenada asociada al perfil que se desea calcular. En la práctica deben realizarse en primer lugar simulaciones de dinámica molecular de cada uno de los estados que se desean conectar a través del perfil de energía libre. Luego de asegurarse que se ha realizado una exploración eficiente de cada uno de los estados (estados inicial y final), se obtienen las estructuras iniciales para el cálculo de MSMD. Se realizan un numero de simulaciones (30-40) partiendo del estado inicial al final eligiendo cuidadosamente la velocidad guía (ν), dado que si se utiliza un valor muy alto se corre el riesgo de sobre calentar el sistema pero si se realiza de forma muy lenta se estará utilizando recursos computacionales de forma poco eficiente. En cada una de las simulaciones se registra el trabajo en función de la coordenada y al finalizar se realiza el promedio exponencial para obtener la energía libre. Dado que la ecuación de Jarzynski es válida solamente si se realiza un número infinito de simulaciones y que esto resulta impracticable, debe realizarse una cantidad suficiente de simulaciones en las que pueda considerarse que se ha logrado la convergencia del sistema, es decir, que la realización de nuevas simulaciones no modifica significativamente el perfil de energía libre final. El método de MSMD se encuentra implementado en AMBER utilizado para realizar los cálculos de dinámica molecular clásica y QM-MM. (Crespo, Martí, Estrin, et al., 2005)

2.4.2 Determinación del ΔG_U mediante MM/PBSA.

Debido a las aproximaciones que realizan las heurísticas de *docking* al estimar la energía libre de unión es razonable realizar estimaciones de ΔG_U con métodos de mayor calidad. Si bien los valores absolutos están muy alejados de valores obtenidos experimentalmente si permiten realizar comparaciones entre compuestos (el $\Delta\Delta G$) (Miller III et al., 2012) Brevemente, el método de MM/PBSA se basa en la descomposición de la energía libre de unión en cuatro componentes:

$$\Delta G^0_{union,solv} = \Delta G^0_{union,vacio} + \Delta G^0_{solv,complejo} - (\Delta G^0_{solv,receptor} + \Delta G^0_{solv,ligando}) \tag{2.15}$$

donde la contribución de solvatación de cada componente se calcula de la siguiente forma utilizando la ecuación de Poisson-Boltzmann linealizada para cada estado (el componente electrostático) agregando un término empírico para la contribución hidrofóbica.

$$\Delta G^0_{solv} = G^0_{electroestatico,\epsilon=80} - G^0_{electroestatico,\epsilon=1} + \Delta G^0_{hidrofobico} \tag{2.16}$$

Mientras que el componente en vacío se calcula tomando un promedio de las interacciones entre el receptor y el ligando (ΔE^0_{MM}) y el cambio entrópico por la unión si es necesario:

$$\Delta G^0_{vacio} = \Delta E^0_{MM} - T\Delta S^0_{modosnormales} \tag{2.17}$$

Mediante el uso de esta técnica de post procesamiento se obtiene un mejor estimador del ΔG_U de las distintas drogas salidas del análisis de *Docking*.

2.5 *Docking* molecular

Las metodologías de docking (o encastre molecular en castellano, me tomaré la licencia y utilizaré el nombre en inglés a lo largo del desarrollo de la tesis) permiten, a partir de un campo de fuerzas simplificado (respecto a uno tipo AMBER), el tratamiento del receptor como un cuerpo rígido y una heurística de búsqueda estimar el modo de unión de una molécula pequeña (un sustrato, una droga, etc) a un receptor (en general, proteico pero puede ser de otra naturaleza) de forma muy rápida. En esta sección se explicaran los dos elementos necesarios para construir un algoritmo de *docking*: i) La función de puntuación, que permite clasificar las distintas poses de una misma molécula como así también comparar entre distintas moléculas; y ii) el algoritmo de búsqueda, que permite a partir de una conformación inicial generar una pose dentro del receptor cuya relevancia debe ser evaluada con la función de puntuación mencionada en i).

Si bien en las anteriores secciones tratamos un método (Dinámica molecular acoplada al análisis de MM-PBSA) que permite estimar de manera precisa la energía libre asociada a la

unión (ΔG) acarrea un costo computacional enorme si se los desea utilizar para estimar la energía libre de unión de bibliotecas de compuestos grandes. Aún si fuera posible desde el punto de vista computacional aparece otro problema, determinar la posición inicial sobre la cual se realizarían los cómputos. Por otra parte, para realizar cálculos de dinámica molecular clásica se necesitan parámetros de unión (distancias de enlace, ángulos, diedros) y de no unión (radios de VdW y cargas parciales). Si bien los primeros se pueden obtener del GAFF en el caso de AMBER, el costo que implica realizar estimaciones de cargas con el método RESP utilizando AM1-bcc resulta prohibitivo para bibliotecas más grandes que unos cientos de moléculas.

Por estas razones se recurre a métodos simplificados de *docking* (o encastre) en las que los hidrógenos no polares se encuentran representados de forma implícita, las cargas parciales se computan con un método aproximado y rápido (Gasteiger, basado en las diferencias de electronegatividad de los átomos, (Gasteiger, & Marsili, 1978)). En el caso de los parámetros de unión como las moléculas se consideran cuerpos semi-rígidos no son necesarios, solo algunos diedros tienen permitida la movilidad y el valor, en general, se trata de una constante fija respecto a la estructura de partida. En el receptor se realizan dos aproximaciones: La primera es tratarlo como un cuerpo rígido, ahorrando una gran cantidad de cómputo; la otra aproximación es que se selecciona una región de la proteína en donde se realiza la simulación de *docking*. Esto permite acelerar enormemente los cálculos generando unas 100 poses para un ligando pequeño (¡5 enlaces rotables) en 30 segundos - 1 minuto. El programa utilizado para realizar la búsqueda virtual de compuestos es rDock (Li, Chen, & Weng, 2003; Ruiz-Carmona et al., 2014)

2.5.1 Función de puntuación

La función de puntuación es una herramienta que permite comparar y clasificar, de un modo veloz, las distintas soluciones propuestas por el algoritmo de docking. Estas funciones de puntuación nacen de la necesidad de poder clasificar una serie de poses de una molécula para poder elegir las ”mejores”, es decir, la que mejor representen el modo de unión de esa molécula X con un receptor Y. Otro rol que cumplen las funciones de puntuación es permitir comparar entre distintas moléculas para saber cual es potencialmente el mejor compuesto que se une a un receptor dado. Existen distintos tipos de funciones de puntuación pudiendo clasificarlas en dos grandes familias: i) las basadas en la física (physics-based), como los campos de fuerza clásicos, y ii)

las empíricas, desarrolladas en base a resultados experimentales, en general, de cristalografía de rayos-X. Es usual que las funciones de puntuación sean mixtas, por ejemplo en el caso de rDock, los parámetros de unión y la electrostática esta basada en el campo de fuerzas Tripos 5.2 mientras que los componentes que se utilizan para estimar el cambio de solvatación son de carácter empírico.

En esta sección se realizará una pequeña descripción de la función de puntuación que utiliza rDock. Una descripción más detallada de los parámetros y ecuaciones se pueden observar en el manual de rDock en la web (`http://rdock.sourceforge.net/wp-content/uploads/2015/08/rDock_User_Guide.pdf`).

La función de puntuación total tiene cuatro componentes: Un puntaje que evalúa la energía de interacción S^{inter}, uno que evaluá la energía interna del ligando S^{intra}, otro que evalúa la energía de cambios en el sitio de unión S^{site} si el receptor se trata con cierta flexibilidad y una que evalúa la energía para todas las restricciones no físicos que se le desee aplicar al sistema (salir fuera de la cavidad definida, RMN, farmacofóricos y de sub-estructura) $S^{restricciones}$

$$S^{total} = S^{inter} + S^{intra} + S^{site} + S^{restraint} \tag{2.18}$$

$$S^{inter} = W^{inter}_{vdw} * S^{inter}_{vdw} + W^{inter}_{polar} * S^{inter}_{polar} + W^{inter}_{repul} * S^{inter}_{repul} + W^{inter}_{arom} * S^{inter}_{arom} + W_{solv} S_{solv} + W_{rot} N_{rot} + W_{const} \tag{2.19}$$

$$S^{intra} = W^{intra}_{vdw} * S^{intra}_{vdw} + W^{intra}_{polar} * S^{intra}_{polar} + W^{intra}_{repul} * S^{intra}_{repul} + W^{intra}_{diedro} * S^{intra}_{diedro} \tag{2.20}$$

$$S^{sitio} = W^{sitio}_{vdw} * S^{sitio}_{vdw} + W^{sitio}_{polar} * S^{sitio}_{polar} + W^{sitio}_{repul} * S^{sitio}_{repul} + W^{sitio}_{diedro} * S^{sitio}_{diedro} \tag{2.21}$$

$$S^{restricciones} = W_{cavidad} * S_{cavidad} + W_{tether} * S_{tether} + W_{RMN} * S_{RMN} + W_{ph4} * S_{ph4} \tag{2.22}$$

El potencial de vdW utilizado en rDock tiene una forma funcional similar al utilizado en el programa de *docking* GOLD (Verdonk et al., 2003). Los tipos de átomos y radios de vdW fueron tomados del campo de fuerza Tripos 5.2 (Clark, Cramer, & Van Opdenbosch, 1989). Dado que Tripos se trata de un campo de fuerza del tipo *all atom* fueron desarrollaros tipos de átomos para los carbonos que contienen hidrógenos de forma implícita (un modelo de representación conocido comúnmente como united-atom). El radio de vdW se incrementa en 0.1 Åpor cada hidrógeno implícito sin cambiar la profundidad de los pozos.

La forma funcional se puede cambiar entre un potencial suave 4-8 y uno más restrictivo 6-12. Un potencial cuadrático se utiliza a corto alcance para evitar penalidades energéticas excesivas para los choques atómicos.

La porción polar se calcula mediante un potencial empírico tipo Bohm para puntuar puentes hidrógeno y otras interacciones polares de corto alcance. Los términos polares se dividen en dos porciones S_{polar} y S_{repul}, que tratan la parte atractiva y repulsiva respectivamente. Seis tipos distintos de centros polares son considerados: donores de puente hidrógeno, iones metálicos, carbonos cargados positivamente (como lo son los carbonos centrales de los grupos guanidinio, amidonio e imidazol), aceptores de puente hidrógeno con una direccionalidad pronunciada por los pares libres, aceptores con una preferencia planar pero sin direccionalidad por pares libres y todo el resto de los aceptores.

El potencial de desolvatación implementado el rDock combina una aproximación basada en la superficie accesible pesada (WSAS, por sus siglas en Inglés) con una aproximación probabilística rápida de la superficie accesible a solvente (SASA) basado en distancias interatómicas de a pares y radios. S_{solv} es computado como el cambio en la energía de solvatación del ligando y del sitio de unión luego de la unión del mismo. Las energías de referencia se toman de las conformaciones iniciales del ligando y del sitio respectivamente y no de la pose siendo evaluada. Se realiza de este modo para tomar en cuenta cualquier cambio en la solvatación de índole intramolecular.

Las energías de los diedros son calculados utilizando los parámetros de Tripos 5.2 para todos los ligandos con las correcciones correspondientes por las contribuciones faltantes por la representación con hidrógenos no polares implicitos.

2.5.2 Algoritmo de búsqueda

En la sección anterior se describió la forma de clasificar las distintas poses pero no como se generan. Dado que los modos de no unión no se encuentran bien definidos es necesario enumerar todos los casos posibles (que son infinitos) y evaluar su puntaje para poder encontrar la solución exacta transformando al problema, desde el punto de las Ciencias de la Computación, en *NP-hard*. Por esta razón es necesario utilizar heúristicas para poder aproximar una solución, inexacta pero lo suficientemente buena que permita comparar entre las distintas soluciones generadas por el mismo algoritmo. Como en el caso de MM-PBSA, los métodos de *docking* generan soluciones razonables que permiten comparar moléculas entre si pero no estimar de forma fehaciente el $DeltaG_U$. Como la función de puntuación se trata de un estimador del $DeltaG_U$, el algoritmo a utilizar lo que debe hacer es minimizar el valor de dicha función. Existen diversas estrategias para lograr esto, algoritmos genéticos, que están basados ligeramente en conceptos de Genética y Evolución dado que hay Darwinianos y Lamarckianos, búsquedas mediante Monte-Carlo (MC), Simulated annealing (SA, enfriamiento simulado en castellano), etc. Se explicaran brevemente dos de esas estrategias, los algoritmos genéticos y las búsquedas mediante MC.

Los algoritmos genéticos se basan ligeramente en las nociones biológicas de genética y evolución. Las propiedades a optimizar se representan en un ”cromosoma” siendo cada una de ellas un ”gen” (por ejemplo, el centro de masa del ligando o el valor de un ángulo diedro). Al comienzo de cada ciclo se genera una población inicial de cromosomas con valores provistos al azar y se les permite ”cruzarse” entre si para generar nuevas soluciones conservándose las de mejor *fitness* (la que tiene el menor valor de puntuación). A estas mejores soluciones se le producen ”mutaciones” (cambios al azar en el valor de alguna de las propiedades a optimizar) y el proceso se repite por otra generación. El algoritmo continua hasta que el valor de puntuación alcance un umbral o bien, se alcance un número máximo de generaciones definido por el usuario.

Por otra parte, los métodos de Monte-Carlo se utilizan para resolver problemas de grados de libertad acoplados, como en el caso del *docking*. Aplicados a métodos moleculares se utiliza para evolucionar un sistema de un estado A a uno B, con una probabilidad basada en Boltzmann para aceptar movimientos, utilizando una cadena de Markov.

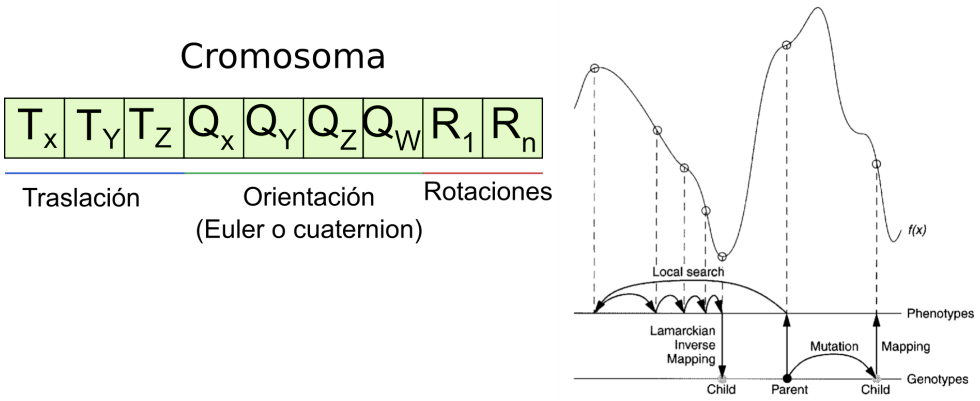


Figura 2.1: **Búsqueda de soluciones de *docking* mediante algoritmos genéticos lamarkianos.** A la izquierda se muestra un esquema del ”cromosoma” y la estructura de datos que representa junto con las variables a optimizar. A la derecha se muestra un gráfico de como es esa búsqueda en el espacio de variables (eje X) en relación con la función de puntuación (eje Y). Esquema de búsqueda tomado de el manual de Autodock 4.

El proceso de búsqueda, análogo a cualquier otro tipo de implementación de heurísticas basadas en algoritmos genéticos, comienza con la construcción de un cromosoma en donde se representan las variables relevantes para definir el modo de unión de un ligando a un receptor rígido: la traslación en X, Y y Z; una forma de definir la orientación respecto al receptor, ángulos de Euler o un cuaternión; y los enlaces rotables (diedros). Se genera una población con valores en cada una de las variables al azar, se entrecruzan dichos cromosomas para generar una población de hijos (*childs*), se evalúa el *fitness* de cada uno de los individuos y se conserva el mejor (el conformero de menor energía) al cual se le realiza una búsqueda local para mejorar el resultado, dicho resultado es el hijo (*child*). El proceso se repite hasta alcanzar un criterio de convergencia o hasta un número finito de interacciones (numero de generaciones). Ver figura 2.1

rDock utiliza una combinación de técnicas de búsqueda estocásticas y deterministas para generar poses de baja energía de los ligandos. El protocolo de *docking* estándar que devuelve una sola pose (un solo resultado) consiste en tres etapas de un algoritmo genético (GA1, GA2 y GA3), seguido de una simulación corta de Monte-Carlo (MC) a baja temperatura y una minimización tipo Simplex (MIN) (Nelder, & Mead, 1965). Las etapas GA son independientes y pensadas para ser utilizadas de forma secuencial. Varios parámetros de puntuación son variados entre cada uno de los GA para promover un muestreo eficaz de las poses iniciales, a su vez minimizando la probabilidad que las poses queden atrapadas temprano en la búsqueda. Las variaciones se presentan en la forma funcional del potencial de VdW (cambiando de un potencial 4-8

en GA1 y GA2 a un potencial 6-12 en GA3, MC y MIN), la tolerancia del potencial polar y las funciones angulares (relajadas en GA1 y progresivamente ajustadas en GA2/GA3/MC) y el peso del potencial de diedro de los ligandos (reducido en GA1 e incrementado en GA2/GA3/MC). Todos los parámetros de la función de puntuación se encuentran en sus valores finales para las etapas de MC/MIN. El cromosoma de GA consiste en el centro de masa del ligando (COM), la orientación del ligando, representada por los ángulos de Euler (Preseción, Nutación y Rotación intrínseca) necesarios para rotar el ligando en su eje principal en el eje cartesiano de referencia, los ángulos diedros rotables del ligando y los ángulos diedros rotables del receptor(si tiene). La población inicial es generada para que el centro de masa del ligando se posicione en algún lugar al azar de la grilla seleccionada y, la orientación y los ángulos diedros del ligando aleatorizados. Las mutaciones son aplicadas al azar a algún grado de libertad y su magnitud elegida a partir de una distribución rectangular de ancho definido.

Una generación se considera completada cuando el numero de nuevos individuos creados es igual al tamaño poblacional. En vez de tener un numero fijo de generaciones, al GA se le permite continuar hasta que la población converge (la mejora del puntaje es menor a 0,1 unidades por las últimas tres generaciones). Esto permite una terminación temprana de poses de bajo rendimiento (puntaje malo) para la cual la población inicial no es capaz de generar una buena solución. Una vez que los pasos de GA han convergido, se realiza una simulación de Monte-Carlo a baja temperatura para refinar la pose seguido de una minimización Simplex para generar una solución minimizada.

2.5.3 RMSD

Existen diversas técnicas para analizar los cambios de un sistema a lo largo de una trayectoria respecto a una referencia, la más utilizada se trata del desvío cuadrático medio que se encuentra definido como:

$$RMSD_j = \sqrt{1/N \sum_{i=1}^{N} (x_{j,i} - x_{ref,i})^2} \tag{2.23}$$

donde N es el número de átomos del sistema, $x_{j,i}$ es la posición del átomo i en la foto j; $x_{ref,i}$ es la posición del átomo i en la foto de referencia. Se pueden realizar cálculos de RMSD a lo largo de una dinámica molecular para observar, de forma global, que le sucede al sistema respecto a la referencia.

Bibliografía

Arcon, Juan Pablo et al. (2015). "Molecular Mechanism of Myoglobin Autoxidation: Insights from Computer Simulations". In: *The Journal of Physical Chemistry B* 119.5, pp. 1802–1813.

Bayly, Christopher I et al. (1993). "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model". In: *The Journal of Physical Chemistry* 97.40, pp. 10269–10280.

Berendsen, Herman JC et al. (1984). "Molecular dynamics with coupling to an external bath". In: *The Journal of chemical physics* 81.8, pp. 3684–3690.

Cieplak, Piotr et al. (1995). "Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins". In: *Journal of Computational Chemistry* 16.11, pp. 1357–1377.

Clark, Matthew, Richard D Cramer, & Nicole Van Opdenbosch (1989). "Validation of the general purpose Tripos 5.2 force field". In: *Journal of Computational Chemistry* 10.8, pp. 982–1012.

Crespo, Alejandro, Marcelo A Martí, Darío A Estrin, et al. (2005). "Multiple-steering QM-MM calculation of the free energy profile in chorismate mutase". In: *Journal of the American Chemical Society* 127.19, pp. 6940–6941.

Crespo, Alejandro, Marcelo A Martí, Susana G Kalko, et al. (2005). "Theoretical study of the truncated hemoglobin HbN: exploring the molecular basis of the NO detoxification mechanism". In: *Journal of the American Chemical Society* 127.12, pp. 4433–4444.

Crespo, Alejandro, Damián A Scherlis, et al. (2003). "A DFT-based QM-MM approach designed for the treatment of large molecular systems: Application to chorismate mutase". In: *The Journal of Physical Chemistry B* 107.49, pp. 13728–13736.

Darden, Tom, Darrin York, & Lee Pedersen (1993). “Particle mesh Ewald: An N log (N) method for Ewald sums in large systems”. In: *The Journal of chemical physics* 98.12, pp. 10089–10092.

Defelipe, Lucas A et al. (2015). “Protein Topology Determines Cysteine Oxidation Fate: The Case of Sulfenyl Amide Formation among Protein Families”. In: *PLoS computational biology* 11.3, e1004051–e1004051.

Dumas, Victoria G et al. (2014). “QM/MM study of the C—C coupling reaction mechanism of CYP121, an essential cytochrome p450 of Mycobacterium tuberculosis”. In: *Proteins: Structure, Function, and Bioinformatics* 82.6, pp. 1004–1021.

Elstner, M (2006). “The SCC-DFTB method and its application to biological systems”. In: *Theoretical Chemistry Accounts* 116.1-3, pp. 316–325.

Elstner, Marcus et al. (1998). “Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties”. In: *Physical Review B* 58.11, p. 7260.

Eschrig, H, & I Bergert (1978). “An optimized LCAO version for band structure calculations application to copper”. In: *physica status solidi (b)* 90.2, pp. 621–628.

Gasteiger, Johann, & Mario Marsili (1978). “A new model for calculating atomic charges in molecules”. In: *Tetrahedron Letters* 19.34, pp. 3181–3184.

Grubmüller, Helmut et al. (1991). “Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions”. In: *Molecular Simulation* 6.1-3, pp. 121–142.

Hornak, Viktor et al. (2006). “Comparison of multiple Amber force fields and development of improved protein backbone parameters”. In: *Proteins: Structure, Function, and Bioinformatics* 65.3, pp. 712–725.

Jarzynski, Christopher (1997). “Nonequilibrium equality for free energy differences”. In: *Physical Review Letters* 78.14, p. 2690.

Li, Li, Rong Chen, & Zhiping Weng (2003). “RDOCK: Refinement of rigid-body protein docking predictions”. In: *Proteins: Structure, Function, and Bioinformatics* 53.3, pp. 693–707.

Martí, Marcelo A et al. (2008). “Mechanism of product release in NO detoxification from Mycobacterium tuberculosis truncated hemoglobin N”. In: *Journal of the American Chemical Society* 130.5, pp. 1688–1693.

Miller III, Bill R et al. (2012). "MMPBSA. py: an efficient program for end-state free energy calculations". In: *Journal of Chemical Theory and Computation* 8.9, pp. 3314–3321.

Nelder, John A, & Roger Mead (1965). "A simplex method for function minimization". In: *The computer journal* 7.4, pp. 308–313.

Porezag, Dirk et al. (1995). "Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon". In: *Physical Review B* 51.19, p. 12947.

Ruiz-Carmona, Sergio et al. (2014). "rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids". In: *PLoS Comput Biol* 10.4, e1003571.

Salomon-Ferrer, Romelia et al. (2013). "Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald". In: *Journal of Chemical Theory and Computation* 9.9, pp. 3878–3888.

SETTLE, Miyamoto S Kollman PA (1992). "An analytical version of the SHAKE and RATTLE algorithm for rigid water molecules". In: *J. Comput. Chem* 13, pp. 952–962.

Turjanski, Adrian Gustavo, Gerhard Hummer, & J Silvio Gutkind (2009). "How mitogen-activated protein kinases recognize and phosphorylate their targets: A QM/MM study". In: *Journal of the American Chemical Society* 131.17, pp. 6141–6148.

Verdonk, Marcel L et al. (2003). "Improved protein–ligand docking using GOLD". In: *Proteins: Structure, Function, and Bioinformatics* 52.4, pp. 609–623.

Wang, Junmei et al. (2004). "Development and testing of a general amber force field". In: *Journal of computational chemistry* 25.9, pp. 1157–1174.

Zhechkov, Lyuben et al. (2005). "An efficient a posteriori treatment for dispersion interaction in density-functional-based tight binding". In: *Journal of Chemical Theory and Computation* 1.5, pp. 841–847.

Capítulo 3

Selección de blancos proteicos basado en criterios de expresión, sensibilidad a estrés y drogabilidad contextual

3.1 Introducción

En pos de contribuir a la búsqueda de nuevos fármacos desde el punto de vista de los blancos, en la presente tesis se ha generado una base de datos de proteínas de todo el genoma de *Mtb.* llamado TuberQ, que relaciona el análisis de la drogabilidad estructural de todas las proteínas de *Mtb.* con estructura depositada en el *Protein Data Bank* (PDB) como así también modelos generados mediante modelado comparativo con las propiedades de unión de drogas de sitios putativos y eventualmente reuniendo información derivada de proteínas similares como así también información sobre la esencialidad, los niveles de expresión en distintas condiciones y un criterio de *off-target*. TuberQ es una base de datos de drogabilidad estructural que contiene todas las estructuras de *Mtb.* resueltas anteriormente y modelos obtenidos por un *pipeline* de modelado comparativo desarrollado en el laboratorio en conjunto con su drogabilidad estructural, esencialidad, la relevancia y un criterio de *off-target*. La combinación de información estructural (drogabilidad) y fisiológica (esencialidad) hacen de TuberQ una herramienta útil para, por ejemplo, descartar genes que aparecen como buenos desde un punto de vista biológico pero que no tienen bolsillos drogables, o en el descubrimiento nuevos bolsillos drogables, incluyendo sitios alostéricos, en blancos ya conocidos. La base de datos permite una inspección simple

y rápida de estructuras proteicas y la drogabilidad de los bolsillos en el contexto de la información experimental disponible teniendo en cuenta la relevancia como blanco terapéutico. Los antibacterianos ejercen su función biológica en una condición fisiológica dada. Para incluir esta propiedad se ha incorporado información relativa a la esencialidad de cada gen-proteína, que cuando es inhibida resultaría en un efecto bacteriostático o bactericida. (Agüero et al., 2008) La esencialidad de Mtb esta basada en ensayos de mutagénesis a escala genómica (Sassetti, Boyd, & Rubin, 2003; Sassetti, & Rubin, 2003) , estudios *in silico* basados en en análisis de flujos metabólicos (Jamshidi, & Palsson, 2007; Raman, Rajagopalan, & Chandra, 2005), las proteínas sensibles a la presencia de NO (a través de sus cisteinas y/o tirosinas) y en la determinación de cuellos de botella metabólicos. (Hasan et al., 2006) La relevancia de los blancos terapéuticos en el estado patológico, varios trabajos en la última década han observado la asociación de genes mediante el uso de microarreglos de DNA en distintas condiciones que imitan aspectos importantes del ambiente que enfrenta el bacilo dentro del macrófago. (Sassetti, & Rubin, 2003; Rengarajan, Bloom, & Rubin, 2005; Voskuil, Bartek, et al., 2011; Betts et al., 2002; Hampshire et al., 2004; Muttucumaru et al., 2004; Boshoff, & C. E. Barry, 2005) TuberQ incorpora información extraída de literatura curada manualmente de esencialidad y expresión bajo condiciones de estrés.

3.2 Materiales y métodos

3.2.1 Armado de de la base de datos TuberQ

El pipeline de TuberQ consiste en los siguientes pasos, descriptos resumidamente en la Figura **??**. Las secuencias de los marcos abiertos de lectura de *Mtb.* (ORFs) y sus meta-datos asociada fueron bajados de la base de datos UniProt (Consortium, 2008). Todos los ORFs son luego analizados con el programa HMMer (Johnson, Eddy, & Portugaly, 2010) y los dominios estructurales asignados. Luego, cada ORF es utilizado para realzar una búsqueda con BLAST contra el Protein Data Bank (PDB) para determinar si la estructura del ORF (o una parte de ella) ha sido resuelta. Basado en estos resultados, cada ORF(o dominio) es clasificado como ‘Resuelto‘ o ‘No resuelto‘. La estructura de los ORFs no resueltos (o dominios) es modelada de acuerdo con nuestro *pipeline* si un molde adecuado esta disponible. Para todas las estructuras,

tanto las experimentales como las modeladas, diversas propiedades estructurales son calculadas incluyendo: (i) la función de puntuación de drogabilidad (Druggability Score DS) para cada bolsillo, (ii) la similitud con proteínas humanas (para evaluar el potencial efecto *off-target*), (iii) los residuos del sitio activo (si están disponibles), (iv) los residuos conservados o relevantes de la familia de PFAM y (v) la potencial sensibilidad a especies reactivas de oxígeno y nitrógeno (ERON) debido a la presencia de residuos específicos o co-factores en el sitio activo. Esta información es luego combinada con los criterios de esencialidad y la información derivada de los experimentos de expresión diferencial en el *pipeline*-motor ProteinQ.

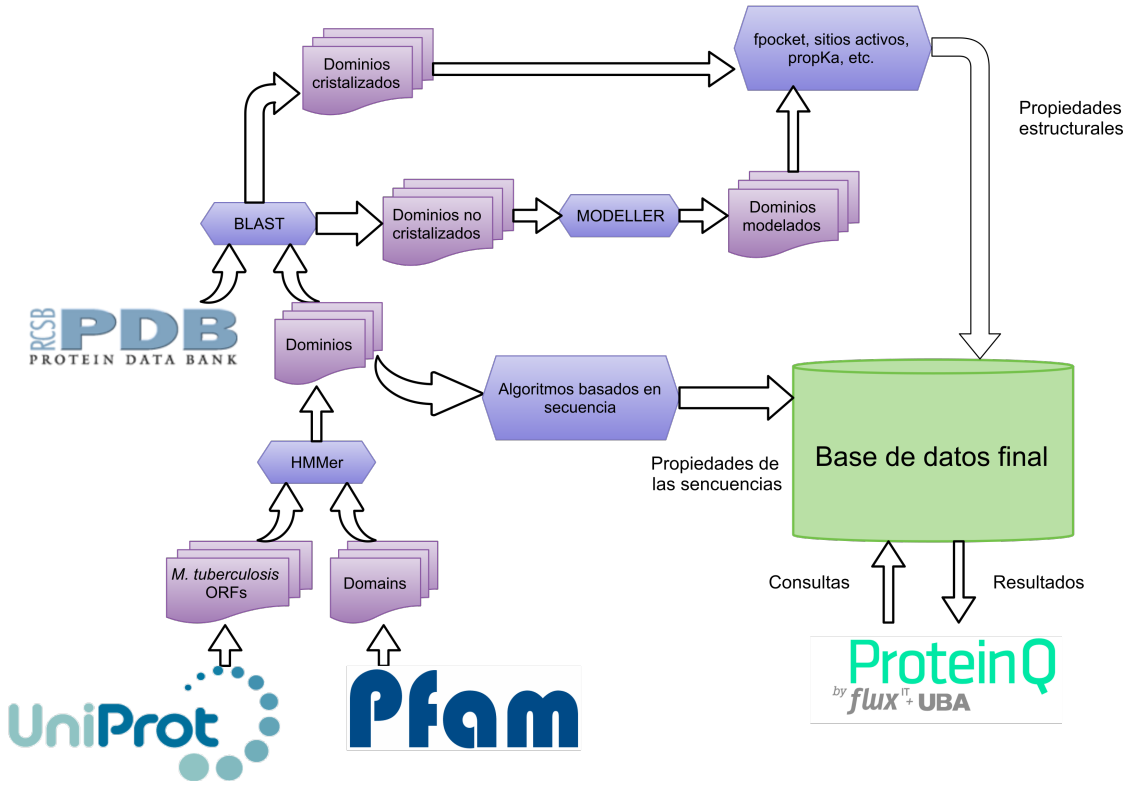


Figura 3.1: **Representación esquemática del *pipeline* de TuberQ.** A partir del genoma traducido (los ORFs) y utilizando el PDB, Modeller, BLAST ,HMMer y PFAM es posible determinar si existen estructuras o si es necesario modelarlas además de determinar el/los dominios que le corresponden a cada ORF. Luego se pueden realizar sobre el estructuroma calculos adicionales como la *bindability* mediante fpocket, la determinación de sitios activos o la sensibilidad a ERON por presencia de Cisteinas y/o Tirosinas.

A continuación brindamos una descripción detallada de los programas y bases de datos utilizadas para cada uno de los pasos del *pipeline*.

3.2.2 Obtención de las secuencias proteicas de *Mtb*.

Todos los ORFs o posibles proteínas de *Mtb* H37rv como fueron derivadas de la secuenciación del genoma entero (Cole et al., 1998) fueron bajas de la base de datos UniProt (www.uniprot.org, código de organismo 3A1773) (Consortium, 2008) En total se obtuvieron 3982 ORFs.

3.2.3 Asignación de dominios/familias PFAM

Todos los ORFs fueron analizados con HMMer (Johnson, Eddy, & Portugaly, 2010) y asignados a una familia o dominio de PFAM, totalizando 5822 asignaciones de dominio a PFAM-A, 1446 dominios a PFAM-B y 1255 ORFS sin dominio asignado. El numero de ORFs con dominio asignado es de 1920, aproximadamente un 48% de los ORFs. Sin embargo, como es de esperar, más de un ORF puede ser asignado a un mismo dominio. Entonces considerado esta información pudimos asignar 1658 dominios únicos (distintos) en todo el genoma de *Mtb.*. En promedio, el genoma de *Mtb.* tiene 2,13 dominios por ORF y 1,19 dominios únicos por ORF.

3.2.4 Selección de la información de expresión por microarreglos

Para determinar que blancos son relevantes en condiciones de estrés, hemos llevado a cabo un análisis combinando de múltiples reportes de expresión génica en microarreglos realizados en una variedad de condiciones que se sospecha dominan el estado de latencia de *Mtb*. Dada la falta de conocimientos de las condiciones fisiológicas reales en la fase de latencia, varios modelos de imitación del ambiente han sido diseñados como son la hipoxia, la falta de nutrientes y el co-cultivo con macrófagos entre otros.(Betts et al., 2002; Hampshire et al., 2004; Muttucumaru et al., 2004; Karakousis et al., 2004; Ohno et al., 2003; Rengarajan, Bloom, & Rubin, 2005; Schnappinger et al., 2003; Talaat et al., 2004; Voskuil, Schnappinger, et al., 2003; Voskuil, Bartek, et al., 2011; Robinson, Adolfsen, & Brynildsen, 2014) Hasta donde sabemos, este es el análisis más completo estudiado y representa una actualización al realizado por Murphy y Brown en 2007. (Murphy, & Brown, 2007)

3.2.5 Criterio de esencialidad

Hemos incluido los cuatro criterios de esencialidad disponibles a escala genómica para *Mtb*. Rubin y colaboradores desarrollaron una serie de estudios utilizando una técnica genética denominada *Transposon Site Hybriziation* (TraSH). Dicha técnica consiste en la inserción al azar de un elemento genético móvil para producir un *knockout* en un gen. (Sassetti, Boyd, & Rubin, 2003). Esta técnica fue utilizada en un estudio de viabilidad *in vitro* (Sassetti, Boyd, & Rubin, 2003), y la biblioteca de mutantes resultantes fue también utilizada en un modelo murino C57BL/6J para determinar la abundancia relativa de las diferentes lineas de *Mtb*.(Sassetti, & Rubin, 2003) De este trabajo unos 192 genes (p-valor <0,005) fueron agregados como esenciales en condiciones *in vivo* en nuestra base de datos. En un tercer estudio, un análisis de supervivencia con macrófagos fue realizado con la misma biblioteca de mutantes TraSH. (Rengarajan, Bloom, & Rubin, 2005) Finalmente, en un cuarto estudio, Sassetti y colaboradores utilizaron una estrategia similar basada en el sistema mutacional del transposon *himar1* para determinar la frecuencia de inserciones brindando una actualización del trabajo realizado por Rubin y colaboradores. (Griffin et al., 2011)

3.2.6 Generación de los modelos estructurales basados en homología

Hasta ahora existen 467 estructuras únicas pertenecientes a *Mtb* depositadas en el PDB. Para el resto de los ORFs intentamos construir modelos basados en homología utilizando el siguiente *pipeline* estructural. Para todos los ORFs de *Mtb*, el primer paso consiste en realizar un PSI-BLAST contra una biblioteca de modelos, que incluye todas las secuencias de cada cadena individual en el PDB agrupadas al 95% de identidad con CD-hit, esto es para reducir la redundancia intrínseca que posee el PDB. (W. Li, & Godzik, 2006) Luego, cada estructura blanco fue creada utilizando MODELLER (Eswar et al., 2008), utilizando el alineamiento local obtenido en el paso de búsqueda por PSI-BLAST. (Altschul et al., 1997) Para cada secuencia blanco, se construyeron 10 modelos distintos y su calidad fue evaluada utilizando los métodos GA341 (Melo, & Sali, 2007) y QMEAN (Benkert, Tosatto, & Schomburg, 2008). Únicamente los modelos con un puntaje de GA341 mayor a 0,7, un QMEAN entre -2 y 2 y una cobertura mayor al 60% fueron utilizados. Este procedimiento brindó unas 2061 estructuras modeladas de alta calidad, que abarcan el 60% de los ORFs de *Mtb*.

3.2.7 Determinación de la drogabilidad estructural

La drogabilidad estructural de cada potencial blanco fue evaluada determinando (y caracterizando) la habilidad de los bolsillos putativos de unir un compuesto tipo droga utilizando fpocket (Le Guilloux, Schmidtke, & Tuffery, 2009) y el recientemente desarrollado DrugScore (DS). (Schmidtke, & Barril, 2010) El método esta basado en el algoritmo de teselación de Voronoi para identificar las cavidades y computar los descriptores fisicoquímicos correspondientes (la densidad hidrofóbica media normalizada, el puntaje hidrofóbico y el puntaje de polaridad normalizado) que se combinan para brindar el DS (las ecuaciones **??**-**??**) Los parámetros de ajuste a las funciones se encuentran publicados en (Schmidtke, & Barril, 2010).

$$DrugScore(x) = \frac{e^{-z}}{1 + e^{-z}} \tag{3.1}$$

$$z = \beta_0 + \beta_1 f_1(d_1) + \beta_2 f_2(d_2) + \beta_3 f_3(d_3) \tag{3.2}$$

$$f_x(d_x) = \frac{e^{-\beta_{x,0}+\beta_{x,1}d_x}}{1 + e^{-\beta_{x,0}+\beta_{x,1}d_x}} \tag{3.3}$$

La densidad hidrofóbica media normalizada. Esta propiedad intenta identificar si el bolsillo en cuestión contiene partes que son ‘bastante hidrofóbicas‘. Por cada esfera R apolar, se computan la cantidad de esferas apolares R vecinas buscando el solapamiento entre esferas R. La sumatoria de todas las esferas apolares R en la vecindad es dividido por el número total de esferas apolares R en el bolsillo. Por último, este estadístico es normalizado comparando con otros bolsillos en la misma proteína. (Schmidtke, & Barril, 2010)

El puntaje hidrofóbico. Este descriptor esta basado en la escala de hidrofobicidad publicada por Monera y otros (Monera et al., 1995). Para todos los residuos presentes en el bolsillo, se calcula y se tiene en cuenta cada residuo una sola vez, es decir que solo se lo tiene en cuenta para un bolsillo.

El puntaje de polaridad normalizado. Cada residuo puede ser dividido en dos categorías de polaridad (1 y 2) (como se encuentra descripto en http://www.info.univ-angers.fr/ gh/Idas/proprietes.htm) El puntaje final de polaridad es la media de los puntajes de polaridad de todos los residuos en el bolsillo. Cada residuo solo se tiene en cuenta una vez.

El puntaje de drogabilidad al tratarse de una función logística, puede adoptar valores entre 0 y 1, siendo 0 (no drogable) y 1 (altamente drogable). Basado en un análisis preliminar de la distribución de DS de todos los bolsillos que albergan un compuesto tipo droga presentes en el PDB (Ver (L. Radusky et al., 2014)) en relación a otros menos drogables o no drogables, los bolsillos se clasifican en cuatro categorías (Figura **??**): (i) no drogables (ND; con un DS menor a 0,2), (ii) pobremente drogables (PD; con un DS entre 0,2 y 0,5), (iii) drogables (D; con un DS entre 0,5 y 0,7) y (iv) altamente drogables (HD; con un DS mayor a 0,7). Para discusión más detallada sobre el método de drogabilidad véase (L. Radusky et al., 2014; Schmidtke, & Barril, 2010).

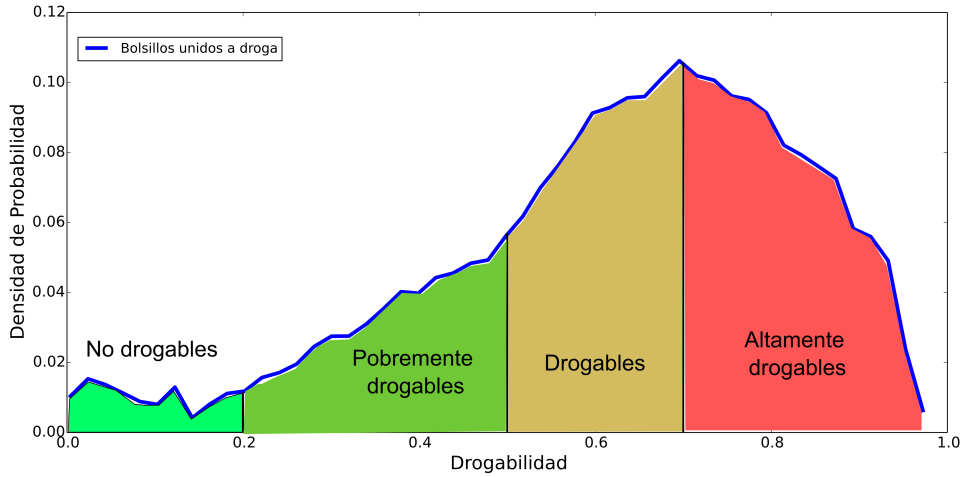


Figura 3.2: **Distribución de bolsillos proteicos en *Mtb.* de acuerdo a su clasificación por DS**. Clasificamos a los bolsillos en cuatro categorías distintas: No drogables si su puntaje se encuentra entre 0 y 0.2, probablemente drogable entre 0.2 y 0.4, drogable si se encuentra entre 0.4 y 0.7 y altamente drogable si se encuentra entre 0.7 y 1.

Tomando en cuenta el estado de oligomerización, para cada proteína cuya estructura ha sido resuelta como un complejo hemos añadido no solamente los cálculos de drogabilidad correspondientes a la sub-unidad sino también la del monómero permitiendo la búsqueda de bolsillos drogables en las interfaces de interacción proteína-proteína permitiendo el desarrollo de drogas contra las mismas. Por último, para tener en cuenta los posibles conflictos relacionados con la flexibilidad proteica, siempre que estén disponibles, hemos computado la drogabilidad estructural de todas las estructuras disponibles para una misma proteína.

3.2.8 Identificación de sitios activos

Para identificar los bolsillos que corresponden al sitio activo y/o determinar la relevancia de un bolsillo en relación a su función, ProteinQ implementa dos análisis distintos basados en: (i) La información depositada en Catalytic Site Atlas (CSA) (Porter, Bartlett, & Thornton, 2004) y (ii) un criterio de importancia basado en PFAM. (Bateman et al., 2004)

La información de CSA (bajada de http://www.ebi.ac.uk/thornton-srv/databases/CSA/) consiste en una lista de identificadores de PDB (PDBId) junto al número de residuo que forman el sitio activo de la proteína. Para mapear el sitio activo de la mayor cantidad posible de dominios, cada PDBId en CSA fue asignado al / a los dominio(s) correspondiente(s). Luego, el consenso del sitio activo fueron transferidos a todas las proteínas de *Mtb.* que tienen ese dominio asignado pero que carecen de entradas en CSA. Esta asignación, basada en el hecho que es esperable que los residuos catalíticos estén conservados en un dominio dado, aproximadamente dobla la cantidad de proteínas cuyos residuos del sitio activo pueden ser identificados.

Como una alternativa para determinar la relevancia de un bolsillo dado (o un residuo), buscamos por residuos en una familia de PFAM/dominio dado que están localizados en una posición importante y estén muy conservados. Las posiciones importantes son aquellas que en el correspondiente modelo de HHMer su contenido de información es mayor a un valor de corte definido (icov). La naturaleza de los aminoácidos conservados en una posición fue determinada comparando la probabilidad de emisión de cada residuo (ep) con icov. Si la relación entre ep e icov era mayor que el valor de corte de un residuo conservado (ctcov), el residuo evaluado se presume conservado. Los valores óptimos de icov y ctcov son de 0,27 y 0,24 respectivamente. Para una descripción más detallada vease (L. Radusky et al., 2014)

Utilizando estos análisis en cada dominio de PFAM, TuberQ brinda una lista de residuos relevantes que pueden ser mapeados en todos los ORFs de *Mtb* con un dominio PFAM asignado.

3.2.9 Evaluación estructural de la sensibilidad a especies reactivas de nitrógeno y oxígeno.

El criterio de sensibilidad a ERON está basado en la combinación de información estructural y de reactividad química. Como fue mencionado anteriormente los principales blancos de ERON son los centros metálicos de las proteínas, como los grupos Hemo, los tioles de las cisteínas y

también tirosinas que pueden ser nitradas. Para las metalo-proteínas la modificación del estado de oxidación/coordinación del centro metálico resulta en una pérdida parcial o total de la actividad, por lo tanto una proteína que contiene centros metálicos con Fe,Cu o Zn será predicha como sensible a ERON si dichos iones son necesarios para su función. La predicción funcional de las modificaciones de cisteínas o tirosinas no es directa pero es un supuesto razonable que si ese residuo tirosina o cisteína es un residuo del sitio activo (o del bolsillo activo), su modificación química puede derivar en una actividad disminuía. Por lo tanto todas las proteínas que tengan un residuo cisteína o tirosina en el sitio activo drogable fueron marcadas como sensibles a ERON.

3.2.10 Construcción de la red metabólica de *Mycobacterium tuberculosis* H37Rv

En pos de construir la red metabólica de *Mycobacterium tuberculosis* se utilizó el programa Pathway Tools and Pathologic, versión 18.0. Pathologic genera una red metabólica (MN) que contiene todos los caminos metabólicos de un organismo dado. En el caso de *Mtb.* H37Rv (GenBank AL123456.3) basado en GenBank y el proteoma en formato FASTA como entradas. El software asocia automáticamente genes con reacciones enzimáticas basado en el número de EC (Enzyme Commission) utilizando la información suministrada por GenBank. Luego de la construcción automática de la red matabólica se procedió a un curado manual de la misma. El curado incluye el borrado de vías mal armadas, el agregado de vías existentes confirmadas por experimentos, o el completado (utilizando la herramienta disponible en Pathway Tools para tal fin). Para determinar números de EC faltantes se realizaron mediante la estrategia de *BLAST bidirectional best hit* en otros genomas bacterianos. Sólo se trabajó con las redes metabólicas que involucran compuestos pequeños filtrando las vías que estén involucradas con ADN, ARN y proteínas.

Luego de la construcción de la red metabólica de *Mycobacterium tuberculosis*, un programa de Python fue escrito para generar una lista con todos los productos y reactivos involucrados en la red y, manualmente inspeccionados para determinar la frecuencia. Dichos compuestos ‘moneda de intercambio‘ como el ATP, co-factores (NADH, FAD, etc) y el agua fueron descartados dado que pueden producir conexiones artificiales en el grafo de la red. Un total de 51

compuestos fueron filtrados antes de la transformación de la red metabólica a un grafo. En pos de identificar potenciales blancos de fármacos realizamos una búsqueda de cuellos de botella (*chokepoints*). Un cuello de botella es una reacción que produce o consume un metabolito único que debe estar balanceado (es decir debe ser producido por una enzima y debe ser consumido por otra). De otra forma se trata de un cuello de botella sin salida (nadie consume el metabolito o nadie lo produce, *Dead End Metabolite* - DEM-). La presencia de DEMs puede reflejar la falta de completitud del la red metabólica en cuestión, por ejemplo falta de reacciones de transporte o metabólicas aunque algunos DEM son auténticos. Por otra parte hemos analizado los datos de las red metabólica de *Mtb.* en el contexto de la información previa de esencialidad, expresión en estrés, drogabilidad y sensibilidad a ERON.

3.3 Resultados

La base de datos TuberQ puede ser accedida y utilizar su interfaz web en http://tuberq.proteinq.com.ar . La interfaz ofrece un menú con varias opciones para obtener la información de la proteína de interés. Estas opciones incluyen el uso de (I) Palabras clave (Nombre de UniProt o cualquier otro criterio; Protein Kinase PknB), (II) UniProtID (Indentificador alfanumérico de UniProtKB; por ejemplo: O05871 para Protein Kinase PknB), (III) PFAMID (Identificador de familia de PFAM; por ejemplo, PF01436.16, NHL repeats) y (IV) PDBID (los cuatro caracteres alfanuméricos de PDB; por ejemplo, 1IDR para *Mtb* Truncated Hemoglobin N). Como ejemplo, asumamos que sabemos el identificador de UniProt de una proteína de interés. En este caso, simplemente ingresamos 'P0A5Y6' en el cuadro de texto y seleccionamos *UniProtID* en el menú desplegable para encontrar todas las entradas asociadas.

Las búsquedas pueden retornar una sola entrada de la base de datos (como cuando se busca por un PDBID o un UniProtID) o múltiples entradas (si se trata de una búsqueda por palabra clave). Los resultados se muestran (ver figura **??**) de acuerdo a su DS de forma ascendente o descendente (Por defecto en forma ascendente). Para cada una de las entradas, el UniProtID, el nombre 'común', el dominio(s) de PFAM y el PDBID o el identificador de modelado por homología son presentados para cada una de las entradas. En el ejemplo que seguimos (inhA), la proteína de interés ha sido cristalizada varias veces y, para cada estructura de Rayos X se puede encontrar el computo de drogabilidad estructural depositado en TuberQ. Al realizar clic

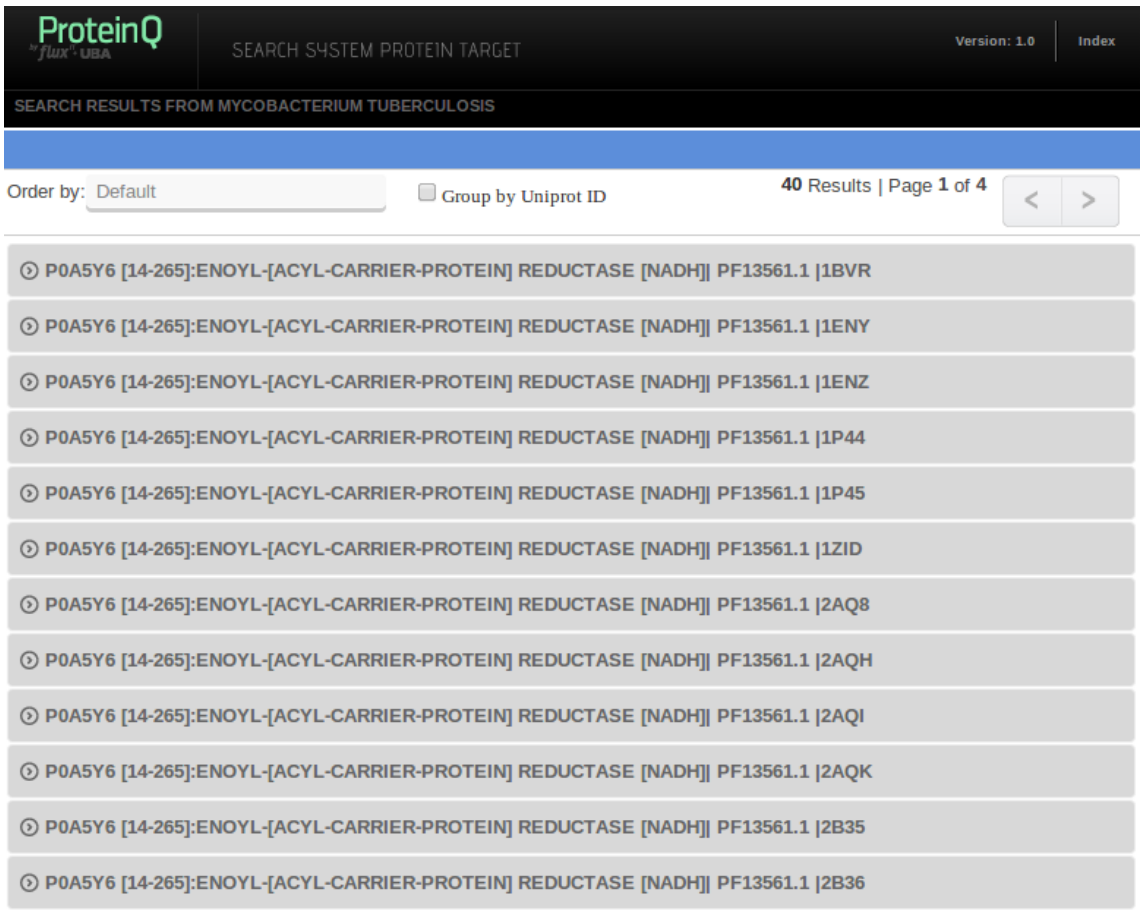
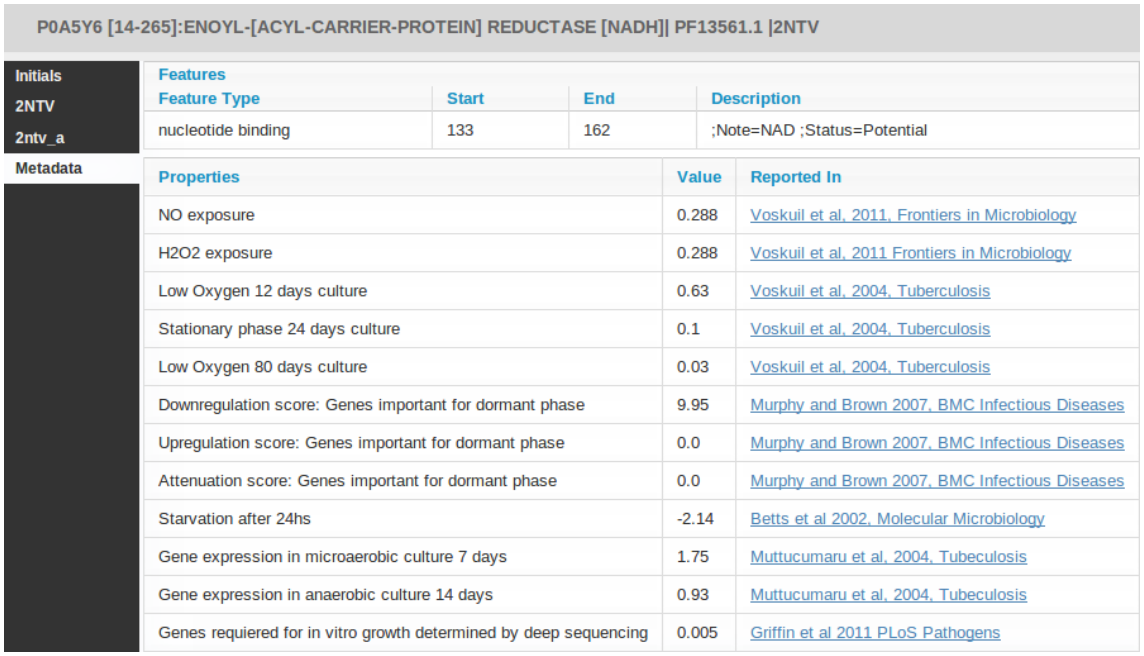


Figura 3.3: **Representación de los resultados de la búsqueda.** Cada triada UniProt-PFAM-Estructura representa una entrada diferente en la base de datos. Se puede elegir agrupar las entradas de UniProt al tildar 'Group by UniProtID'

sobre la entrada, esta se expande brindando más información.



P0A5Y6 [14-265]:ENOYL-[ACYL-CARRIER-PROTEIN] REDUCTASE [NADH]| PF13561.1 |2NTV

Initials
2NTV
2ntv_a
Metadata

Features

Feature Type	Start	End	Description
nucleotide binding	133	162	;Note=NAD ;Status=Potential

Properties	Value	Reported In
NO exposure	0.288	Voskuil et al, 2011, Frontiers in Microbiology
H2O2 exposure	0.288	Voskuil et al, 2011 Frontiers in Microbiology
Low Oxygen 12 days culture	0.63	Voskuil et al, 2004, Tuberculosis
Stationary phase 24 days culture	0.1	Voskuil et al, 2004, Tuberculosis
Low Oxygen 80 days culture	0.03	Voskuil et al, 2004, Tuberculosis
Downregulation score: Genes important for dormant phase	9.95	Murphy and Brown 2007, BMC Infectious Diseases
Upregulation score: Genes important for dormant phase	0.0	Murphy and Brown 2007, BMC Infectious Diseases
Attenuation score: Genes important for dormant phase	0.0	Murphy and Brown 2007, BMC Infectious Diseases
Starvation after 24hs	-2.14	Betts et al 2002, Molecular Microbiology
Gene expression in microaerobic culture 7 days	1.75	Muttucumaru et al, 2004, Tubeculosis
Gene expression in anaerobic culture 14 days	0.93	Muttucumaru et al, 2004, Tubeculosis
Genes requiered for in vitro growth determined by deep sequencing	0.005	Griffin et al 2011 PLoS Pathogens

Figura 3.4: **Solapa *Summary*.** En la solapa *Summary* se encuentra la información de asignación a UniProt, PFAM y PDB y los correspondientes links hacia esas bases de datos; la información acerca de la asignación de dominio y determinación de estructura realizados por el programa HMMer y BLAST respectivamente y el alineamiento entre la proteína de *Mtb* y su homologo más cercano en el genoma humano.

Para cada una de la entradas aparecen 3 solapas principales (siempre accesibles en la parte izquierda de la pantalla). En la solapa 'Initials' (ver figura **??**) información general de la proteína es presentada, en conjunto con la asignación (y los correspondientes enlaces a bases de datos externas) a la familia de PFAM y la estructura (PDB). InhA esta asociada casi en todo su largo a la familia PF1356, que corresponde al dominio 'Enoyl (Acyl Carrier Protein) Reductase'. Para nuestro ejemplo, elegiremos la estructura correspondiente al PDBID 2NV6. Por otro lado, en la solapa 'Initials' se muestra el mejor resultado de realizar BLAST contra el genoma humano.

Seleccionando cualquiera de las estructuras, al realizar clic en el PDBID al costado izquierdo de la pantalla, se presenta la información estructural, incluyendo la visualización interactiva del bolsillo (ver figura **??**). El módulo de visualización permite al usuario (I) seleccionar un bolsillo para visualizarlo, (II) mostrar los HETATOMS y residuos asignados por CSA o PFAM, (III) mostrar la proteína como *cartoon* o como esferas y varillas y (IV) mostrar los residuos que forman parte del bolsillo o sus correspondiente *alpha spheres*. En el ejemplo, mostramos las *alpha spheres* del *pocket* '0' en verde, dado que es un bolsillo altamente drogable (HD),

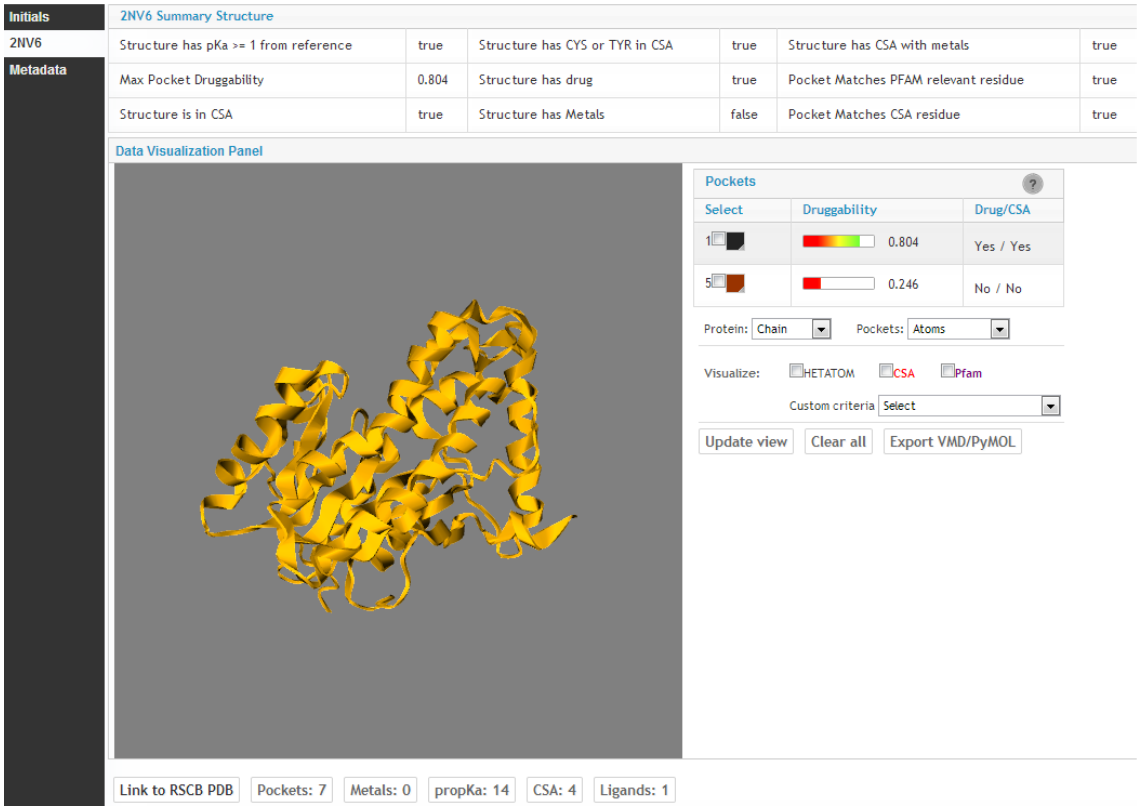


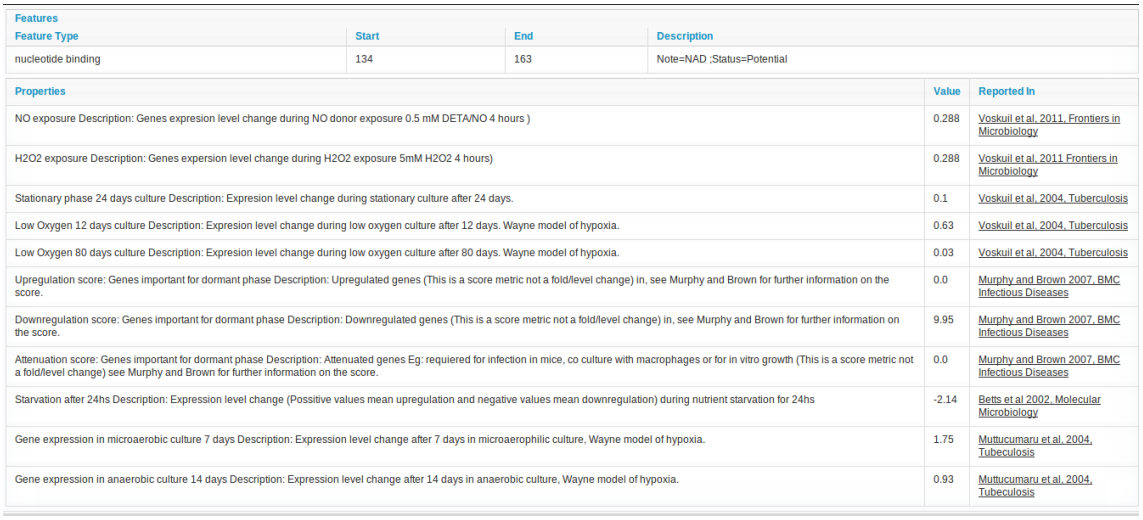
Figura 3.5: **Solapa *Structure*.** En en el panel superior se muestra la información sobre la drogabilidad máxima, la presencia de drogas o los residuos reportados en CSA. En el panel central se encuentra el visualizador de estructuras. El panel derecho permite al usuario controlar el visualizador (decidir que mostrar y como mostrarlo). Debajo se presenta, mayor información sobre cada entrada, como es la cantidad total de bolsillos encontrados, los residuos con desvíos en su pKa en solución (realizado con el programa propKa) o la presencia de metales.

los HETATOMS encontrados en el cristal como esferas y la proteína como cintas. Otra visualización posible del mismo bolsillo incluye a los residuos que definen el bolsillo (en vez de las *alpha spheres*, ver figura **??**) y los residuos reportados como parte del sitio activo para ver si alguno coincide con el bolsillo drogable. La visualización puede realizarse también en VMD (Humphrey, Dalke, & Schulten, 1996) o PyMol (DeLano, 2002) al bajar el archivo comprimido correspondiente.

Figura 3.6: **Solapa *Structure* 2.** El bolsillo es mostrado como un conjunto de *alpha spheres* (polares-verdes- y apolares -blancas-), mientras que el ligando se encuentra representado por el tipo de átomo (carbonos, gris; nitrógenos, azul; oxígeno, rojo; azufre, amarillo y fósforo, violeta). En este caso el ligando corresponde a la droga isoniazida unida a NAD.

Información adicional es provista en la parte inferior de la solapa *Structure*. Por ejemplo, detalles de los ligandos que han sido co-cristalizados (ZID en el caso de 2NV6) pueden ser obtenidos. Información completa de todos los bolsillos hallados en la proteína por el programa fpocket es también accesible realizando click en el botón 'pockets' en la parte inferior de la página. La correspondiente pagina muestra todos los bolsillos hallados ordenados por su *Drug Score* (como fue definido en los métodos) como así también otros parametros como su volumen, número de esferas alfa mientras que solamente los bolsillos que han sido clasificados como HD o D se muestran en la solapa '*Structure*' aquí se muestran todos. Por último, en la última solapa, 'Metadata' la información de otras bases de datos es mostrada (UniProt principalmente)

como así también los datos recopilados manualmente sobre expresión en distintas condiciones que imitan la infección (Estrés Nitrosativo, Estrés Oxidativo, Hipoxia, escasez de nutrientes y perfiles de expresión durante la infección en modelos murinos). (Ver figura **??**.



Features

Feature Type	Start	End	Description
nucleotide binding	134	163	Note=NAD ;Status=Potential

Properties	Value	Reported In
NO exposure Description: Genes expresion level change during NO donor exposure 0.5 mM DETA/NO 4 hours)	0.288	Voskuil et al, 2011, Frontiers in Microbiology
H2O2 exposure Description: Genes expersion level change during H2O2 exposure 5mM H2O2 4 hours)	0.288	Voskuil et al, 2011 Frontiers in Microbiology
Stationary phase 24 days culture Description: Expresion level change during stationary culture after 24 days.	0.1	Voskuil et al, 2004, Tuberculosis
Low Oxygen 12 days culture Description: Expresion level change during low oxygen culture after 12 days. Wayne model of hypoxia.	0.63	Voskuil et al, 2004, Tuberculosis
Low Oxygen 80 days culture Description: Expresion level change during low oxygen culture after 80 days. Wayne model of hypoxia.	0.03	Voskuil et al, 2004, Tuberculosis
Upregulation score: Genes important for dormant phase Description: Upregulated genes (This is a score metric not a fold/level change) in, see Murphy and Brown for further information on the score.	0.0	Murphy and Brown 2007, BMC Infectious Diseases
Downregulation score: Genes important for dormant phase Description: Downregulated genes (This is a score metric not a fold/level change) in, see Murphy and Brown for further information on the score.	9.95	Murphy and Brown 2007, BMC Infectious Diseases
Attenuation score: Genes important for dormant phase Description: Attenuated genes Eg: requiered for infection in mice, co culture with macrophages or for in vitro growth (This is a score metric not a fold/level change) see Murphy and Brown for further information on the score.	0.0	Murphy and Brown 2007, BMC Infectious Diseases
Starvation after 24hs Description: Expression level change (Possitive values mean upregulation and negative values mean downregulation) during nutrient starvation for 24hs	-2.14	Betts et al 2002, Molecular Microbiology
Gene expression in microaerobic culture 7 days Description: Expression level change after 7 days in microaerophilic culture, Wayne model of hypoxia.	1.75	Muttucumaru et al, 2004, Tubeculosis
Gene expression in anaerobic culture 14 days Description: Expression level change after 14 days in anaerobic culture, Wayne model of hypoxia.	0.93	Muttucumaru et al, 2004, Tubeculosis

Figura 3.7: **Solapa *Metadata*.** La solapa *Metadata* permite visualizar la información de anotación funcional de UniProt como así también los perfiles de expresión génica en una diversidad de condiciones experimentales como son la exposición a ERON, hambruna, hipoxia y la infección de murinos. En el caso de la esencialidad si tiene un valor de ”1” o de ”YES” el gen/proteína es esencial. Para el caso de los experimentos de sobre expresión en diversas condiciones la interpretación resulta un poco más complicada, valores mayores a 2 se consideran que el gen en cuestión se encuentra sobre expresado mientras que valores menores a 0,5 se considera que el gen se encuentra reprimido en dicha condición.

3.3.1 Estadísticas de TuberQ

La construcción de TuberQ nos permitió analizar algunos datos estadísticos interesantes sobre la drogabilidad del genoma de *Mtb* H37Rv. Desde una perspectiva puramente estructural (de un total de 1344 estructuras, que incluyen tanto a las resueltas por difracción de rayos-X como a los modelos por homología, representando un 34% de los ORFs), el 82% corresponde a proteínas con bolsillos altamente drogables (DS > 0,7). Este hallazgo es alentador para los proyectos de diseño de fármacos pero seguramente refleja también el sesgo en la determinación de estructuras con ligando unido (es decir, estructuralmente drogables) en el PDB. Es importante señalar que una proteína posea un bolsillo drogable es un condición necesaria pero no suficiente dado que la unión a dicho bolsillo debe además modificar la actividad biológica de la proteína en el sentido deseado. Además, generalmente la evaluación de relevancia de un determinado bolsillo debe realizarse de manera manual, dado que el efecto biológico puede involucrar bolsillos

más allá del sitio activo (como pueden ser sitios alostéricos y de interacción proteína-proteína). En este contexto, TuberQ ofrece una forma de inspeccionar fácilmente el bolsillo en conjunto con información acerca de los residuos del sitio activo, residuos relevantes según la familia de PFAM o en el contexto de interacciones proteína-proteína como así también la información de la esencialidad para el crecimiento de la bacteria. Al combinar los criterios de esencialidad y drogabilidad. unos 379 genes (un 9,5 % de todos los ORFs) resultan necesarios para el crecimiento de *Mtb*, unos 352 pueden ser identificados como drogables resultando en un 8,8 % del genoma y un 26% del estructuroma. De este conjunto, 184 proteínas son considerados altamente drogables (HD) (un 4,6% del genoma y un 13% del estructuroma). Finalmente, si se considera la información acerca de sobre expresión bajo condiciones de estrés, que involucran 713 ORFs, 145 son esenciales, 475 son HD y 111 satisfacen todos los criterios. En las siguientes secciones se discutirá en más detalle la utilización de otros criterios para realizar una priorización de blancos.

3.3.2 Clasificación del estructuroma de *Mtb.* por su esencialidad y *bindablity*

Comenzamos nuestro análisis clasificando todos los dominios con estructura disponible (incluyendo aquellos derivados de difracción de rayos-X como los modelos por homología) de acuerdo a su drogabilidad estructural. Para ello dividimos los dominios en cuatro grupos. El primer grupo corresponde al control positivo, es decir, proteínas de *Mtb* que ya han sido cristalizadas con compuestos tipo droga, grupo que llamaremos 'Cristalizados con droga o (CWD)'. El segundo grupo lo denominaremos 'Drogables por extensión al dominio (DDE)', que incluye a todas aquellas proteínas si al menos existe una estructura depositada en el PDB con una droga o compuesto tipo droga en la familia de PFAM respectiva. Entonces el grupo DDE incluye a todas las proteínas que pueden ser drogables. Finalmente, de acuerdo con un criterio de asociación por dominios, el grupo 'resto (R)' contiene todas las estructuras que no tienen relación con ninguna estructura con compuestos tipo droga. Este conjunto se subdivide entre las estructuras de Rayos X (RWC) y los modelos por homología (RWM). Para todas las estructuras computamos todos los posibles bolsillos y el correspondiente puntaje de drogabilidad (Druggability score - DS-) utilizando fpocket (Schmidtke, & Barril, 2010). Primero se analizaron todos

aquellos bolsillos que contienen compuestos tipo droga y como era de esperar representaban los bolsillos con mayor DS. Para el grupo DDE seleccionamos el bolsillo que coincide con el bolsillo de la proteína que ha sido cristalizada con droga que pertenece al mismo dominio. Finalmente, para el grupo Resto analizamos los bolsillos que coinciden con la predicción de sitio activo depositada en CSA, la realizada a través de los residuos importantes de PFAM o ambos cuando estuviera disponible. Habiendo seleccionado los bolsillos relevantes, clasificamos a cada grupo de ORFs en cuatro categorías respecto a su DS. Estas categorías son las arriba mencionadas Non Druggable (ND), Possibly Druggable (PD), Druggable (D) y Highly Druggable (HD). Los resultados están mostrados en la Tabla **??** debajo.

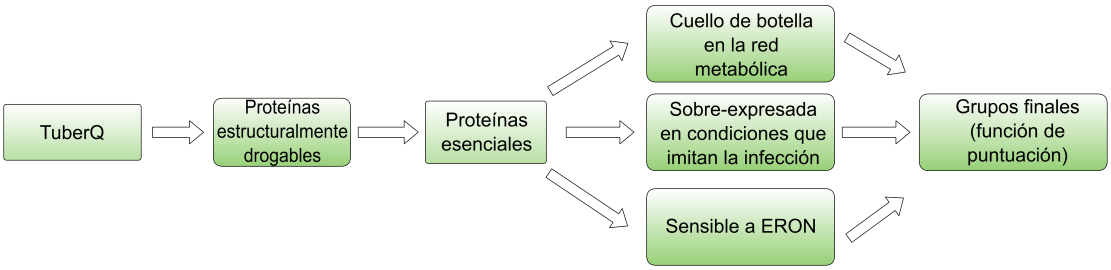


Figura 3.8: ***Pipeline* de clasificación utilizado**. Para determinar la relevancia de una proteína (y su vía metabólica correspondiente) se utilizó la información depositada originalmente en TuberQ agregando los cómputos de vías metabólicas (proteínas que son cuellos de botella), expresión en condiciones que imitan la infección y de sensibilidad a ERON. Estos datos se encuentran actualmente depositados en TuberQ.

Puntaje/Grupo	**CWD**	**DDE**	**RWC**	**RWM**	**Total**
ND	1 (0)	5 (2)	1 (0)	1(0)	8(2)
PD	15 (9)	23 (14)	7 (3)	16 (2)	61 (30)
D	34 (20)	82 (45)	14 (8)	68 (20)	198 (93)
HD	75 (42)	187 (100)	52 (20)	321 (99)	635 (261)
Total	125 (71)	297 (161)	74 (31)	406 (122)	902 (385)

Tabla 3.1: **Proteínas de *Mtb* clasificadas de acuerdo a su *Druggability Score* (DS).** Los números entre paréntesis indican el número de proteínas que son esenciales como se definió en la sección métodos

Los resultados muestran que, como era de esperar, la mayor parte de las proteínas de *Mtb* cristalizadas en presencia de un compuesto tipo droga tienen un DS alto, perteneciendo al grupo HD. En este grupo aparecen proteínas como *enoyl-ACP reductase InhA* (P9WGR1), que es el blanco primordial de la droga de primera linea para el tratamiento de TB, isoniazida, como así también *Hydroxymycolate synthase mmaA4* (Q79FX8) y *Serine/threonine-protein kinase Pkn B* (P9WI81) ambas proteínas para las cuales existen inhibidores (S-adenosyl-N-decyl y Ser/Thr-

mitoxantrone) que han demostrado tener efectos bacteriostáticos (Wehenkel et al., 2006). Por lo tanto y en consonancia con trabajos previos (Schmidtke, & Barril, 2010; L. Radusky et al., 2014) nuestro método es capaz de predecir con un alto grado de certeza la probabilidad de una proteína de albergar un compuesto tipo droga en uno de sus bolsillos. Desde un punto de vista general, lo que resulta interesante es que casi la mitad de las estructuras analizadas (tanto cristalográficas como modelos) son propensas a unir un compuesto tipo droga, este valor es más grande que el computado utilizando sólo un análisis basado en dominios (Hopkins, & Groom, 2002) de cerca de 21% y probablemente refleje el sesgo hacia la determinación de estructuras que ya se sabe son blanco de drogas.

El primer grupo de interés, donde nuevos blanco pueden ser encontrados, abarca a las proteínas dentro del grupo DDE-HD, el echo que tanto el criterio de asociación (ser asignado a DDE) como el criterio estructural (El DS) coinciden para muchos casos, es un fuerte argumento para la selección de 187 proteínas, de las cuales 100 han sido reportadas como esenciales durante el crecimiento *in vitro* y por lo tanto interesantes para un análisis posterior. También, hay cerca de 360 proteínas entre los cristales y los modelos, de los cuales 119 son esenciales que han sido predichas como drogables desde un punto de vista puramente estructural. Una lista completa de los resultados son presentadas en el material suplementario de (Defelipe et al., 2015) y serán analizadas en mayor profundidad al integrar los datos de expresión en la siguiente sección. Las estructuras y sus bolsillos se encuentran disponibles en linea en TuberQ (http://tuberq.proteinq.com.ar/).

3.3.3 Priorización de proteínas de *Mtb* de acuerdo a su perfil de expresión en condiciones tipo infección

Para continuar clasificando las 200 proteínas (altamente) drogables y esenciales identificadas arriba como los mejores candidatos, realizamos un análisis con la información disponibles acerca de los niveles de expresión durante condiciones que imitan la infección. Las condiciones seleccionadas, que agrupan diferentes trabajos, comprende hipoxia, escasez de nutrientes, estrés de ERON e infección en modelos murinos. Primero clasificamos todas las proteínas de acuerdo al número de condiciones en las que se encuentran sobre-expresadas, por lo tanto una proteína con un *Expression Score* (ES) de 0 no se encuentra sobre-expresada en ninguna condición, mientras

que una proteína con un ES de 4 se encuentra sobre-expresada en todas las condiciones aquí evaluadas. (Esta información se encuentra para cada proteína en la Tabla Suplementarias de (Defelipe et al., 2015)) La Tabla **??** muestra una visión global de las proteínas sobre-expresadas en *Mtb*.

ESb	**DDE**	**RWC**	**RWM**	**Total**
4 or 3	17 (7)	5 (1)	16 (5)	38 (13)
2 or 1	192 (66)	50 (14)	282 (61)	524 (141)
0	88 (27)	19 (5)	108 (33)	215 (65)

Tabla 3.2: **Número de proteínas sobre-expresadas en (1 a 4) condiciones tipo infección-** a) Números entre paréntesis corresponden solamente a proteínas drogables y esenciales b) El *Expression Score* (ES) describe el número de condiciones donde la proteína fue encontrada en sobre-expresión, desde 0 (la proteína no se sobre-expresa en condiciones tipo infección) a 4 (la proteína se sobre-expresa en las cuatro condiciones, hipoxia, hambruna, Estrés ERON e infección en ratones).

La tabla **??** muestra que hay 38 proteínas drogrables, 13 de las cuales además resultan esenciales, que están sobre-expresadas en 4 o 3 de las condiciones tipo infección. En el grupo DDE encontramos, por ejemplo, proteínas como la*Redox sensor histidine kinase response regulator DevS* (P9WGK3), conocida por estar involucrada en la transducción de señales en presencia de ERON que contiene un dominio kinasa, con un bolsillo de unión a ATP. Un caso más interesante resulta la *3-methyl-2-oxobutanoate hydroxymethyltransferase* (o *Ketopantoate hydroxymethyltransferase KPHMT*, UniProtID P9WIL7), una proteína que ha sido involucrada tanto en hipoxia como infeccion. Entre el grupo Resto, encontramos proteinas como por ejemplo *L,D-transpeptidase 2* (UniProtID I6Y9J2), la *Alpha-beta hydrolase* (UniProtID I6XU97) y la *DNApol III delta subunit* (UniProtID O06363). Es interesante señalar que la mayoría de las proteínas 'esenciales' están sobre expresadas en 1 o 2 condiciones, una observación que posiblemente refleje el hecho que la expresión proteica de *Mtb* este altamente regulada y adaptable al sutil cambio de condiciones o estímulos externos y muestra que las proteínas sobre-expresadas realizan funciones clave.

3.3.4 Incorporación de un criterio de sensibilidad a estrés de ERON

Como se mencionó anteriormente, una hipótesis para combatir TB es identificar que proteínas ya son blanco de las especies reactivas de nitrógeno y oxígeno (ERON) producidas por el sistema inmune del hospedador e intentar inhibirlas también de forma farmacológica. Por lo tanto,

y además del análisis de expresión, utilizamos la información de estructura-secuencia combinada con el conocimiento de la reactividad química para predecir la sensibilidad de las mismas frente a las ERON. Como fue descripto anteriormente, el principal blanco de estas especies son los centros metálicos de las proteínas, como son el grupo hemo, y los residuos de cisteína y tirosina que pueden ser nitrados/oxidados. Usualmente la modificación del estado de oxidación/coordinación de los centros metálicos de las metalo proteínas resulta en una pérdida parcial o total de función (aunque reversible), como ha sido descripto en las P450 de *Mtb* (Ouellet et al., 2009). En el caso de las tirosinas y cisteínas, es una asunción razonable que si estos residuos se encuentran presentes en el sitio activo (o bolsillo), su modificación química puede derivar en una actividad disminuida. Este es el caso de las Cistein Proteasas que se transforman en inactivas al oxidarse la cisteína del sitio activo (J. Li et al., 1997) o en la MnSOD donde la nitración de la tirosina bloquea el sitio de unión del sustrato (Radi, 2004). Con esto en mente, asignamos como potencialmente sensibles a estrés de ERON todas las proteínas que tienen un centro metálico- (Cu, Fe y Zn) adyacente al bolsillo del sitio activo, o un residuo de cisteína/tirosina en el sitio activo y/o putativo de unión a droga. La condición de sensibilidad a estrés para las proteínas descriptas en las secciones anteriores es presentada en las tablas suplementarias de (Defelipe et al., 2015) mientras que el análisis global se presenta en la tabla **??**.

Caracteristica	**Cristal (modelos)**	**Esencial (E)**	**HD[a] y E**	**HD,E y Over[b]**
AS Metal	149	86	57	41
AS Cys	130 (164)	64 (49)	37 (37)	30 (28)
AS Tyr	269 (274)	135 (84)	82 (69)	58 (42)

Tabla 3.3: **Proteínas de *Mtb* predichas como sensibles a ERON.** En a) proteínas altamente drogables b) Proteínas que están sobre expresadas si el puntaje de expresión es mayor a 3. Los números entre paréntesis corresponde únicamente a las proteínas drogables y esenciales.

La información presentada en la Tabla **??** muestra que hay cerca de 800 proteínas que son potencialmente sensibles a ERON debido a la presencia de un átomo metálico, una tirosina o a una cisteína en el sitio activo. La presencia de tirosina es el doble de común que la cisteína o los iones metálicos pero tiene que tenerse en cuenta que su poder predictivo como estimador de la sensibilidad a ERON es moderado. Este resultado puede ser producto de la abundancia relativa diferencial entre tirosinas y cisteínas o su costo metabólico (Krick et al., 2014). La

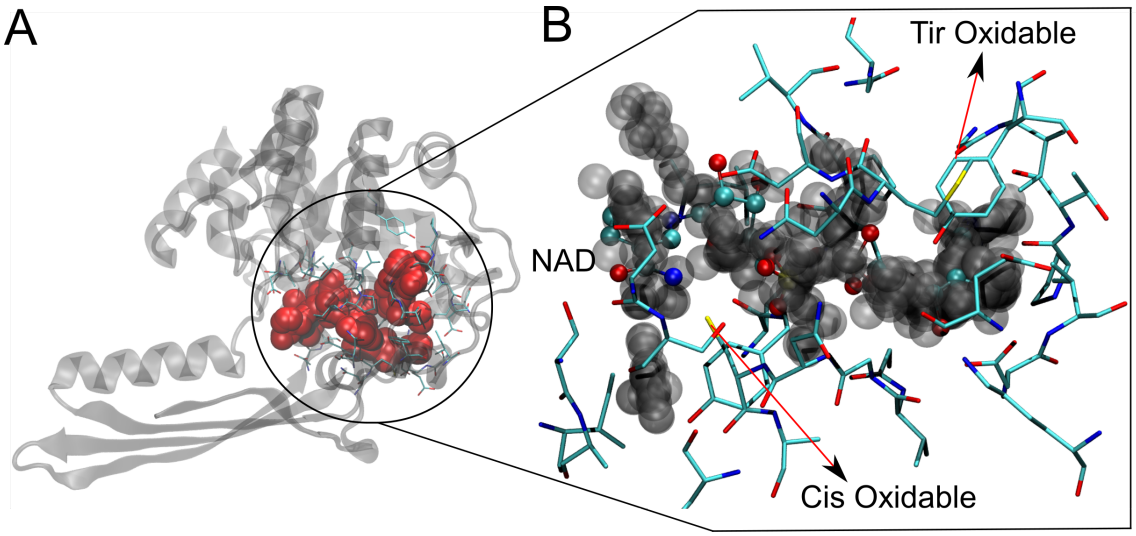


Figura 3.9: **Propiedades estructurales de *Inositol-3-Phosphate Synthase*** A) Vista del plegado de I3PS con el bolsillo drogable destacado en esferas rojas (PDBID 1GR0). B) Acercamiento del bolsillo drogable superpuesto con la estructura de NAD. Las cisteínas y tirosinas oxidables se encuentran dibujadas más gruesas.

combinatoria de los criterios arriba descriptos con el presentado en este apartado indica que hay cerca de 200 proteínas que cumplen todos ellos y por lo tanto se encuentran por arriba en el *ranking*. Más allá del análisis global, es interesante observar cuales proteínas cumplen todos los criterios (Drogable, Esencial, sobre-expresión en ERON) y emergen como posibles blancos de este análisis. Tres casos que llamaron nuestra atención son *Inositol-3-phosphate synthase* (I3PS, ino1) (UniProtID P71703), *L,D-transpeptidase 1* (UniProtID O53638) y el *AraC family transcriptional regulator* (UniProtID P96245). Estas proteínas serán descriptas en la discusión como ejemplos paradigmáticos de los resultados que es posible obtener con este tipo de análisis integrados.

3.3.5 Construcción e incorporación del análisis de redes metabólicas de *Mtb* para priorizar blancos

Cómo último paso en nuestro proceso de priorización, utilizamos el programa Pathway Tools y un curado manual para construir una red metabólica (MN) de *Mtb* y analizamos la unicidad (que sea un cuello de botella) y la centralidad de las reacciones predichas en *Mtb*. Como en los apartados anteriores, la información completa de las MN se encuentra disponible en el sitio de TuberQ. Las características globales de la red se presentan en la tabla **??** y en la figura **??**.

En total se asignaron 985 genes/proteínas únicas correspondientes a 1369 reacciones en-

Característica	Número
Nodos	1569
Bordes	6394
Diámetro direccional	80
Vías	257
Reacciones enzimáticas totales	1708
Reacciones enzimáticas asociadas a un rv	1708
Rv únicos asociados a reacciones enzimáticas	985
Cuellos de botella	509
Rv escenciales	1305

Tabla 3.4: Propiedades generales de la red metabólica de *Mtb*.

zimáticas que se encuentran agrupados en 257 vías distintas. De un total de 1708 reacciones enzimáticas, 1305 están asociadas a genes previamente reportados como esenciales. No pudimos asignar de forma inequívoca unas 339 reacciones que forman parte de de la red metabólica, posiblemente por la falta de caracterización de genes/proteínas, la presencia de reacciones espontaneas y las limitaciones propias de los algoritmos utilizados para reconstruir la red metabólica. Determinamos que 509 reacciones corresponden a cuellos de botella (reacciones que tienen solamente un sustrato o un producto). De estos, un 77% están asociados a genes esenciales, comparado a 55% si se consideran todas las reacciones, por lo tanto, como es de esperar existe una gran coincidencia entre los genes esenciales y los cuellos de botella. Para priorizar los genes/proteínas primero decidimos puntuar cada vía de acuerdo a su relevancia metabólica, determinada por el número de cuellos de botella y su centralidad, como así también su relevancia en condiciones que imitan el estado infectivo utilizando el ES desarrollado en la sección anterior. Los resultados se presentan en la tablas suplementarias de (Defelipe et al., 2015).

El análisis de MN revela varias vías con alta puntuación, por ejemplo la ya conocida vía responsable de la síntesis de micotiol. El micotiol es crucial para mantener el estado redox de la célula regulado y juega un rol pivotal en la supervivencia en macrófagos. (Newton, & Fahey, 2002) Todas las enzimas de la vía son esenciales, el 50% de las proteínas cristalizadas tienen un DS > 0,7 y varias de ellas se encuentran sobre-expresadas en condiciones de estrés de ERON, hipoxia y escasez de nutrientes. Otra vía altamente puntuada es la responsable de la síntesis de histidina, que ha sido sugerida como potencial blanco de fármacos por su ausencia en mamíferos. (Lunardi et al., 2013) Esta vía, compuesta por ocho proteínas que son esenciales y dos proteínas de las cuales se dispone de información estructural, por Rayos X o mediante

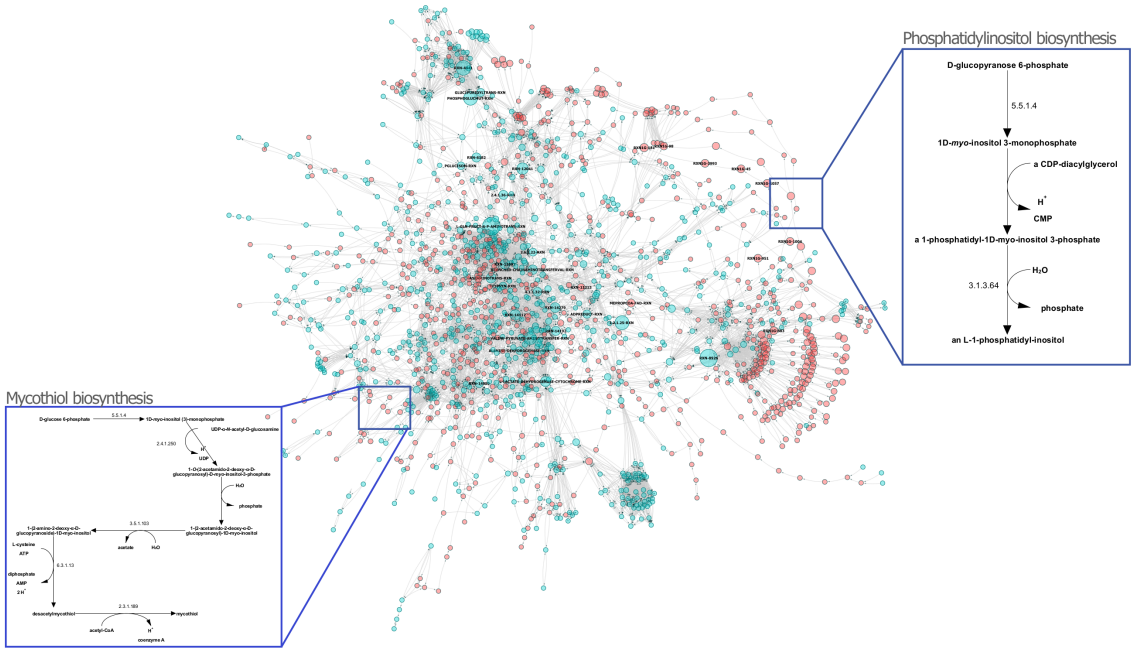


Figura 3.10: **Grafo de reacciones de la red metabólica de *Mycobacterium tuberculosis*.** Cada nodo representa una reacción predicha en el metabolismo de *Mtb*, y existe una arista entre nodos si el producto de una reacción es el sustrato de otra. El tamaño de los nodos representa el valor de *betweenness centrality* en el grafo de reacciones y puntuado primero de acuerdo con esta metrica. Los nodos en rojo representan cuellos de botella. Se destacan dos vías, *mycothiol biosynthesis* y *phosphatidyl-inositol biosynthesis* ambas teniendo altos puntajes según nuestro análisis.

modelado comparativo, y son drogables (DS > 0,59).

Otra vía altamente puntuada es la bien caracterizada vía de síntesis de ácido micólico. El micolato es un componente integral de la pared celular de *Mycobacterium tuberculosis* y participa en la capacidad del bacilo de sobrevivir en el huésped infectado, de virulencia y evasión del sistema inmune. Esta vía es el blanco de las drogas de primera linea isoniasida y etambutol. (C. E. Barry, Crick, & McNeil, 2007) La vía esta compuesto por 22 genes (Rv3804c Rv0470c Rv0242c Rv1483 Rv1483 Rv0242c Rv3720 Rv2524c Rv0974c Rv2247 Rv2502c Rv3280 Rv3799c Rv2524c Rv3720 Rv0636 Rv2245 Rv2246 Rv0644c Rv3372 Rv3801c Rv3800c). La importancia de esta vía yace en el número de genes involucrados de los cuales un 83% son esenciales y un 60% drogables. También un 90% de los genes involucrados resultan ser cuellos de botella.

El análisis de MN también revela la relevancia del metabolismo de azufre cuya importancia (esencialidad) para la supervivencia y la virulencia en muchos patógenos (entre ellos *Mtb*) ha sido demostrada. Es más, la mayor parte de estos genes no se encuentran en humanos. Entre estas vías, el metabolismo de metionina, incluyendo la degradacion a homocisteína, es llevado

a cabo por proteínas drogables (Rv3340 sobre-expresada en tres condiciones, y Rv3341). Particularmente interesante entre las proteínas involucradas es sahH (Rv3248c), cuyo rol en la regulación de L-homocisteina ha sido recientemente reportado (Singhal et al., 2013). Es más, también ha sido implicado como intermediario en mecanismos de resistencia junto con MetK (Rv1392) (Raman, & Chandra, 2008). SahH cataliza la reacción de hidrólisis de SAH a homocisteína y adenosina utilizando NAD+ como co factor. La proteína presenta un plegado alfa/beta compuesto por el motivo de unión a nucleósidos Rossman. (Rao, & Rossmann, 1973) El bolsillo drogable esta delimitado principalmente por residuos polares y cargados negativamente y tiene un volumen de 2284 $Å^3$, bastante grande dado que tiene que acomodar tanto una molécula de NAD+ como de SAH. Tiene dos tirosinas (Tyr 493 y 495) en el sitio de unión a NAD+ convirtiéndola potencialmente en sensible a estrés de ERON. SahH ha sido descripta como esencial en los dos estudios masivos de los que se cuenta con información. (Sassetti, Boyd, & Rubin, 2003; Griffin et al., 2011)

Entre otras vías pequeñas que aparecen bien puntuadas en nuestro sistema de clasificación son aquellas relacionadas con la biosíntesis de lipoato. Los dos genes importantes (Rv2218 [lipA] y Rv2217 [lipB]) son esenciales. LipB es drogable desde un punto de vista estructural y se sobre-expresa bajo escasez de nutrientes. El bolsillo drogable de LipB contiene tres residuos sensibles a ERON Cys 176, Tyr22 y Tyr 91. A pesar de no ser un proceso ubicuo en bacterias, el lipoato ha sido implicado en la patogénesis de las micobacterias incluyendo la respuesta a especies reactivas de oxígeno y nitrógeno producidas por el sistema inmune, también es reconocido que las proteínas lipoliadas juegan un rol en el proceso antioxidante (Allary et al., 2007; Bryk et al., 2002; Spalding, & Prigge, 2010), y por lo tanto, convierten a esta vía en un blanco atractivo desde un punto de vista del desarrollo de fármacos. Por otra parte, LipB ha sido cristalizado y con propiedades terapéuticas promisorias (Ma et al., 2006). Finalmente, otra vía pequeña que aparece en nuestro análisis de MN es *alanine degradation IV* que es realizada por un solo gen (Rv2780).A pesar de no haber sido descripto como esencial, esta *L-alanine dehydrogenase* es el primer antígeno que se encuentra en *M. tuberculosis* pero no en la cepa para vacunación *Mycobacterium bovis* BCG (K. Chan et al., 2002). Adicionalmente, ha sido sugerido que la falta de *L-alanine dehydrogenase* es la razón de la falta de infectividad de *Mycobacterium bovis BCG* en humanos. Esta proteína es drogable y parece cumplir un rol es-

tratégico en la respuesta a estrés nitrosativo dado que se encuentra sobre-expresada en todas las condiciones.

En resumen, nuestro análisis integrado de la red metabólica de *Mtb* en conjunto con la información de expresión, esencialidad y drogabilidad permite la identificación de vías importantes que contienen blancos terapéuticos prometedores. Es más, como será discutido en la siguiente sección toda esta información se encuentra disponible en el sitio web TuberQ y presenta una forma directa y rápida para evaluar el potencial de una proteína como blanco de fármacos para combatir TB.

3.4 Discusión

Dado el potencial de los métodos de análisis a escala genómica, desde la secuenciación del genoma de *Mtb* (Cole et al., 1998), varios trabajos basados en metodologías *in silico* han aparecido sobre el tema (Hasan et al., 2006; Raman, Rajagopalan, & Chandra, 2005; Agüero et al., 2008; Raman, Yeturu, & Chandra, 2008; Jamshidi, & Palsson, 2007). También considerando alguna predicción de drogabilidad (Anand, & Chandra, 2014) y el rol en la fase de latencia basado en perfiles de expresión génica, estos trabajos usualmente incluyen un criterio de esencialidad y de *off-target*. El criterio de esencialidad se refiere a los genes que son esenciales para el crecimiento y/o la supervivencia de la bacteria y por lo tanto, cuando son inhibidos resultarán en un efecto bacteriostático o bactericida. (Agüero et al., 2008). La esencialidad esta basada en estudios experimentales de mutagénesis (Griffin et al., 2011; Sassetti, Boyd, & Rubin, 2003) o en estrategias *in-silico* de análisis de balance de flujos metabólicos (Jamshidi, & Palsson, 2007; Raman, Rajagopalan, & Chandra, 2005) o la determinación de cuellos de botella en la red metabólica (Hasan et al., 2006). Las desventajas de la utilización de este tipo de criterios por eliminar potenciales blancos ha sido recientemente resaltada en relación con el concepto de polifarmacología (Hopkins, 2008). Los estudios genómicos han revelado que debido a la redundancia y a efectos compensatorios solamente un 10-15% de los genes son individualmente esenciales. (Zambrowicz, & Sands, 2004; Winzeler et al., 1999; Giaever et al., 2002) pero que muchos más son 'sintéticamente letales' cuando son *knocked down* en combinación (Hillenmeyer et al., 2008). En este contexto, en este trabajo de tesis, priorizamos la drogabilidad del blanco y la sensibilidad a estrés de ERON y buscamos su rol en el contexto del metabolismo de

Mtb, al destacar vías enteras en vez de proteínas individuales.

Para evitar efectos secundarios adversos se diseñó un criterio de *off-target* dado que es importante notar que el efecto anti-TB debe ser específico para la bacteria y no interferir con las proteínas del hospedador.Esto generalmente se traduce en comparaciones de secuencia (Hasan et al., 2006) o de tipo estructural (Raman, & Chandra, 2008) entre la proteína blanco del huésped y el proteoma del hospedador, y todas aquellas que son demasiado similares al a las del hospedador son descartadas. La dificultad con esta estrategia es que una simple sustitución aminoacídica puede resultar en una unión diferencial, como lo demuestra la aparición de resistencia a antibióticos debido a este tipo de mutaciones y la selectividad de muchas drogas (Hopkins, 2008; C. E. Barry, & Blanchard, 2010). Un caso paradigmático son las diarylquinolinas que inhiben la actividad de la subunidad F0 de la ATP sintetasa (presente en todos los organismos) en la membrana de *Mtb* pero que sin embargo muestra un espectro de actividad muy acotado, perdiendo potencia incluso contra otras actinobacterias, siendo inactiva contra las bacterias gram-positivas y negativas, y unas 20.000 veces más selectiva contra *Mtb* que con su contraparte en mamíferos (Koul et al., 2007; Haagsma et al., 2009). Por lo tanto, creemos que el criterio de *off-target* debe ser muy riguroso y debe tenerse en cuenta luego, en las etapas de desarrollo farmacológico, usualmente en el contexto de estudios estructurales de unión ligando-proteína.

En cualquier caso es interesante comparar estos resultados con los obtenidos por otros *pipelines* de selección de blancos, para ver que hay en común como que nuevos blancos surgen. Muchas proteínas ya largamente nombradas en la literatura como lo son las involucradas en señalización (pknB, pknG, devS), síntesis de ácido micólico (IhnA, pcaA, pks13, fas, fad32D), de pantenoato (panB) y citocromos (cyp121 y cyp125) ya han sido reportados por el laboratorio de Chandra y en TDRTargets (Agüero et al., 2008; Raman, Yeturu, & Chandra, 2008; Anand, & Chandra, 2014). Por otra parte, ninguno de estos reportes destaca proteínas pertenecientes a la vía de síntesis del micotiol (como son mshB e ino1 mencionadas arriba) relevantes para mantener el balance redox en las micobacterias y sugerida como vía relevante para matar al patógeno.(ver figura **??**) La *Inositol-3-phosphate synthase* (I3PS, ino1) (UniProtID P71703) es un miembro de la vía de síntesis del micotiol. Convierte Glucosa-6-P en 1D-myo-inositol-3-fosfato, ha sido descripta como esencial tanto por experimentos masivos (Sassetti, Boyd, &

Rubin, 2003; Griffin et al., 2011) como de mutación de la misma.(Movahedzadeh et al., 2004). Forma parte del regulón DosR y esta sobre-expresada en condiciones de falta de nutrientes. Como se puede observar en la figura **??** la estructura de I3PS (PDBID 1GR0) presenta un bolsillo drogable (DS de 0,719) que se solapa con el sitio de unión a NAD, un sitio conocido por poder albergar compuestos tipo droga en otras proteínas como inhA. Es interesante destacar que I3PS posee dos residuos sensibles a estrés Tyr145 y Cys26 como también un átomo de zinc estructural/catalítico (su rol no es bien comprendido). Claramente I3PS presenta todas las características de un blanco ideal. Otros casos interesantes son la vía de síntesis de lipoato (lipA y lipB) responsables de la síntesis del cofactor enzimático descripto arriba (Ma et al., 2006), y *L-D transpeptidase* que está involucrada en realizar el *crosslinking* de peptidoglicano en la pared celular de *Mtb* (y por lo tanto relacionada con la síntesis de ácido micólico) fundamental para la resistencia *in vivo*. *L-D transpeptidase 1* es la enzima involucrada en la formación de los enlaces entrecruzados del peptidoglicano y por lo tanto esencial. Se encuentra sobre-expresada en todas las condiciones, particularmente en presencia de ERON. El bolsillo drogable (DS de 0,701) que también es el bolsillo del sitio activo contiene la Cisteína 226 que es el nucleófico que actúa en la reacción enzimática, argumentando fuertemente en su inhibición por ERON. Interesantemente, ha sido sugerido también que la actividad catalítica de LDTP1 puede ser inhibida por compuestos beta-lactamicos (Dubée et al., 2012; Cordillot et al., 2013).

3.4.1 Comparación con otros recursos enfocados en drogabilidad

En la década pasada, varios métodos computacionales han sido desarrollados para determinar la drogabilidad de una proteína. (Barril, 2013) La mayor parte de ellos están basados en algoritmos de detección de cavidades para identificar los bolsillos, y utilizan varios descriptores fisico/químicos para realizar sus predicciones. El programa fpocket utilizado en TuberQ pertenece a este grupo. La principal diferencia entre los distintos predictores es el conjunto de estructuras drogables (D) y no drogables (ND) utilizadas para entrenar el método (por ejemplo, utilizar únicamente las estructuras con ligandos que se administran por vía oral) y el subconjunto específico de descriptores de los bolsillos utilizados para entrenar al modelo. La tendencia general muestra que la mayor parte de ellos llegan a un buen nivel de poder predictivo, con tasas de éxito para los casos positivos que rondan el 70-90% (Schmidtke, & Barril, 2010; Sheridan

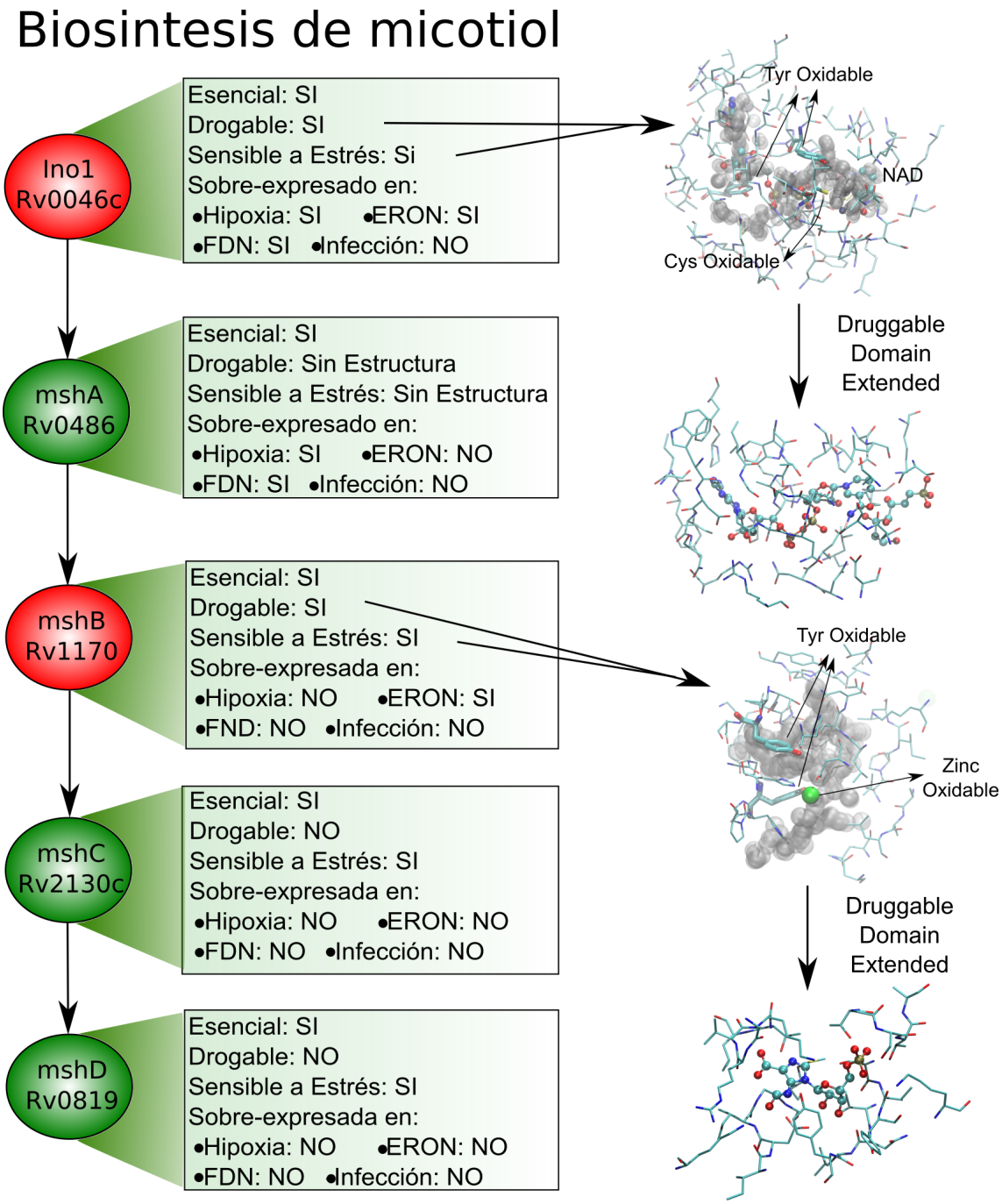


Figura 3.11: **Via de síntesis del micotiol.** En rojo se muestran las proteínas más relevantes de la vía *ino1* y *mshB*. A un lado se muestra una representación del bolsillo drogable y de bolsillo se sabe drogable a través de la asociación por culpa al pertenecer a la misma familia de PFAM. FDN: falta de nutrientes

et al., 2010; Krasowski et al., 2011; Henrich et al., 2010; Volkamer, Griewel, et al., 2010; Pérot et al., 2010; Volkamer, Kuhn, Grombacher, et al., 2012; Desaphy et al., 2012; Perola, Herman, & Weiss, 2012). Es importante notar que, sin embargo, la mayor parte de estos métodos están basados solamente en la estructura y la identificación de las propiedades de los bolsillos, los positivos son indicativos de *bindability* más que de la drogabilidad. Además, la mayor parte de ellos necesita ser bajado, instalado y ejecutado de forma local por el investigador para un blanco, o grupo de blancos, dado. Por lo tanto requiere alguna expertise para obtener la predicción. Hasta donde sabemos, únicamente el método DoGSiteScorer ha sido publicado mediante un servicio Web (Volkamer, Kuhn, Rippmann, et al., 2012).

En este contexto, TuberQ toma como ventaja el uso de métodos de predicción de drogabilidad estructural (fpocket) y brinda información acerca de la drogabilidad clasificando los bolsillos en cuatro simples categorías, que podría simplificar al usuario la interpretación del DS. Además, TuberQ combina los resultados con *metadata* biológica que permite la evaluación directa del potencial impacto terapéutico del blanco. Por otra parte, la información está ya computada y directamente disponible para el investigador (incluso puede ser bajada), transformando al presente recurso, hasta donde sabemos, en único en lo señalado. Es necesario notar que nuestro *pipeline* de modelado por homología a escala genómica permitió la inclusión de más de 900 nuevas estructuras, que pueden ser visualizadas y comparadas con las estructuras de rayos-X disponibles y permite al usuario evaluar proteínas para las cuales no había información estructural disponible.

Nuestra base de datos ha sido diseñada para ofrecer estas características, porque fue concebida como una herramienta para ayudar en el proceso de decisión en el desarrollo de drogas para *Mtb* de forma interactiva en un marco regularmente actualizado. TuberQ ofrece una gran variedad de aplicaciones. Por ejemplo, la búsqueda de *bindability* en nuestra base de datos puede ayudar a decidir sobre la idoneidad de blanco proteico, o alternativamente un investigador puede estar interesado en buscar información sobre la funcionalidad de una proteína específica y encontrar la metadata asociada con la *bindability* y la localización de bolsillos para la elección de los blancos más prometedores. Finalmente, desde el punto de vista de la red metabólica, nuevamente aparecen varios blancos conocidos como son la vía de síntesis de ácido micólico, relevante para la modulación del sistema inmune y como mecanismo de defensa. Es más, vías

involucradas en algún grado a la respuesta a estrés de ERON se encuentran puntuadas muy favorablemente como son la biosíntesis de NAD, homocisteína o de clusters de Hierro-Azufre. En la tabla **??** se presentan un conjunto de 7 nuevos blancos identificados por nuestro *pipeline* como así también un reaseguro de 6 blancos ya descriptos con sus características computadas.

Protein Name	Rv	Status
Inositol-3-phosphate synthase	Rv0046c	Nuevo blanco
3-phosphoshikimate 1-carboxyvinyltransferase	Rv3227	Nuevo blanco
O-acetylhomoserine aminocarboxypropyltransferase	Rv3340	Nuevo blanco
3-oxoacyl-[acyl-carrier-protein] synthase 2	Rv2246	Nuevo blanco
Octanoyltransferase	rv2217	Nuevo blanco
Bifunctional protein GlmU	Rv1018c	Nuevo blanco
Rv1465	Rv1465	Nuevo blanco
MshB (GlcNAc-Ins deacetylase)	RV1170	Revalidado
Sulfate adenylyltransferase subunit 2	Rv1285	Revalidado
dTDP-glucose 4,6-dehydratase	Rv3464	Revalidado
Enoyl-[acyl-carrier-protein] reductase [NADH]	Rv1484	Revalidado
3-methyl-2-oxobutanoate hydroxymethyltransferase	Rv2225	Revalidado
Mycocyclosin synthase	Rv2276	Revalidado

Tabla 3.5: Blancos nuevos y revalidados encontrados en *Mtb* utilizando la drogabilidad estructural, el análisis de importancia metabólica y la información de expresión en condiciones de estrés de acuerdo a la priorización de TuberQ.

3.4.2 Conclusiones y perspectivas

En este capítulo hemos combinado la mayor cantidad de información relacionada con la sensibilidad, esencialidad, relevancia de las proteínas de *Mtb* con la predicción de drogabilidad estructural y análisis en una base de datos amigable, con facilidades gráficas para la visualización y manipulación estructural. Creemos que esta base de datos puede resultar muy importante para la gente trabajando en el campo del descubrimiento de fármacos, selección de blancos terapeútico y biología estructural de TB. TuberQ es la primera base de datos que provee un análisis comprehensivo de las estructuras de *Mtb* e identificación de bolsillos utilizando un DS. En nuestra base de datos, los usuarios pueden fácilmente encontrar si un blanco deseado, elegido por relevancia, tiene un bolsillo drogable y por lo tanto vale la pena seguir el proceso de desarrollo de fármacos. Planeamos extender el presente análisis para incluir información acerca de las bases moleculares de las cepas MDR y XDR, su potencial relación con la drogabilidad, información de drogas de TB de otras bases de datos, como el *TB Drugome database* (Kinnings

et al., 2010) e información relacionada con la variación génica en TB como Tbvar (Joshi, Dhiman, & Scaria, 2014). Finalmente, creemos que nuestra base de datos presenta características interesante desde un punto de vista bioinformático, dado que hay pocas bases de datos que combinen información de drogabilidad estructural con información funcional y fisiológica a escala genómica. Por último, el *pipeline* de computo de drogabilidad estructural aquí descripto puede ser, y ya ha sido y será extendido a otros patógenos (L. G. Radusky et al., 2015), poniendo especial énfasis en los que causan las llamadas enfermedades desatendidas.

Bibliografía

Agüero, Fernán et al. (2008). "Genomic-scale prioritization of drug targets: the TDR Targets database". In: *Nature Reviews Drug Discovery* 7.11, pp. 900–907.

Allary, Marina et al. (2007). "Scavenging of the cofactor lipoate is essential for the survival of the malaria parasite Plasmodium falciparum". In: *Molecular microbiology* 63.5, pp. 1331–1344.

Altschul, Stephen F et al. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs". In: *Nucleic acids research* 25.17, pp. 3389–3402.

Anand, Praveen, & Nagasuma Chandra (2014). "Characterizing the pocketome of Mycobacterium tuberculosis and application in rationalizing polypharmacological target selection". In: *Scientific reports* 4.

Barril, Xavier (2013). "Druggability predictions: methods, limitations, and applications". In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 3.4, pp. 327–338.

Barry, Clifton E, & John S Blanchard (2010). "The chemical biology of new drugs in the development for tuberculosis". In: *Current opinion in chemical biology* 14.4, pp. 456–466.

Barry, Clifton E, Dean C Crick, & Michael R McNeil (2007). "Targeting the formation of the cell wall core of M. tuberculosis". In: *Infectious Disorders-Drug Targets (Formerly Current Drug Targets-Infectious Disorders)* 7.2, pp. 182–202.

Bateman, Alex et al. (2004). "The Pfam protein families database". In: *Nucleic acids research* 32.suppl 1, pp. D138–D141.

Benkert, Pascal, Silvio CE Tosatto, & Dietmar Schomburg (2008). "QMEAN: A comprehensive scoring function for model quality assessment". In: *Proteins: Structure, Function, and Bioinformatics* 71.1, pp. 261–277.

Betts, Joanna C et al. (2002). “Evaluation of a nutrient starvation model of Mycobacterium tuberculosis persistence by gene and protein expression profiling”. In: *Molecular microbiology* 43.3, pp. 717–731.

Boshoff, Helena IM, & Clifton E Barry (2005). “Tuberculosis—metabolism and respiration in the absence of growth”. In: *Nature Reviews Microbiology* 3.1, pp. 70–80.

Bryk, R et al. (2002). “Metabolic enzymes of mycobacteria linked to antioxidant defense by a thioredoxin-like protein”. In: *Science* 295.5557, pp. 1073–1077.

Chan, Kaman et al. (2002). “Complex pattern of Mycobacterium marinum gene expression during long-term granulomatous infection”. In: *Proceedings of the National Academy of Sciences* 99.6, pp. 3920–3925.

Cole, STea et al. (1998). “Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence”. In: *Nature* 393.6685, pp. 537–544.

Consortium, UniProt et al. (2008). “The universal protein resource (UniProt)”. In: *Nucleic acids research* 36.suppl 1, pp. D190–D195.

Cordillot, Mathilde et al. (2013). “In vitro cross-linking of Mycobacterium tuberculosis peptidoglycan by l, d-transpeptidases and inactivation of these enzymes by carbapenems”. In: *Antimicrobial agents and chemotherapy* 57.12, pp. 5940–5945.

Defelipe, Lucas A et al. (2015). “A whole genome bioinformatic approach to determine potential latent phase specific targets in Mycobacterium tuberculosis”. In: *Tuberculosis*.

DeLano, Warren L (2002). “The PyMOL molecular graphics system”. In:

Desaphy, Jérémy et al. (2012). “Comparison and druggability prediction of protein–ligand binding sites from pharmacophore-annotated cavity shapes”. In: *Journal of chemical information and modeling* 52.8, pp. 2287–2299.

Dubée, Vincent et al. (2012). “Inactivation of Mycobacterium tuberculosis L, D-transpeptidase LdtMt1 by carbapenems and cephalosporins”. In: *Antimicrobial agents and chemotherapy* 56.8, pp. 4189–4195.

Eswar, Narayanan et al. (2008). “Protein structure modeling with MODELLER”. In: *Structural Proteomics*. Springer, pp. 145–159.

Giaever, Guri et al. (2002). “Functional profiling of the Saccharomyces cerevisiae genome”. In: *nature* 418.6896, pp. 387–391.

Griffin, Jennifer E et al. (2011). “High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism”. In: *PLoS pathogens* 7.9, e1002251.

Haagsma, Anna C et al. (2009). “Selectivity of TMC207 towards mycobacterial ATP synthase compared with that towards the eukaryotic homologue”. In: *Antimicrobial agents and chemotherapy* 53.3, pp. 1290–1292.

Hampshire, Tobias et al. (2004). “Stationary phase gene expression of¡ i¿ Mycobacterium tuberculosis¡/i¿ following a progressive nutrient depletion: a model for persistent organisms?” In: *Tuberculosis* 84.3, pp. 228–238.

Hasan, Samiul et al. (2006). “Prioritizing genomic drug targets in pathogens: application to Mycobacterium tuberculosis”. In: *PLoS Computational Biology* 2.6, e61.

Henrich, Stefan et al. (2010). “Computational approaches to identifying and characterizing protein binding sites for ligand design”. In: *Journal of Molecular Recognition* 23.2, pp. 209–219.

Hillenmeyer, Maureen E et al. (2008). “The chemical genomic portrait of yeast: uncovering a phenotype for all genes”. In: *Science* 320.5874, pp. 362–365.

Hopkins, Andrew L (2008). “Network pharmacology: the next paradigm in drug discovery”. In: *Nature chemical biology* 4.11, pp. 682–690.

Hopkins, Andrew L, & Colin R Groom (2002). “The druggable genome”. In: *Nature reviews Drug discovery* 1.9, pp. 727–730.

Humphrey, William, Andrew Dalke, & Klaus Schulten (1996). “VMD: visual molecular dynamics”. In: *Journal of molecular graphics* 14.1, pp. 33–38.

Jamshidi, Neema, & Bernhard Ø Palsson (2007). “Investigating the metabolic capabilities of Mycobacterium tuberculosis H37Rv using the in silico strain iNJ661 and proposing alternative drug targets”. In: *BMC systems biology* 1.1, p. 26.

Johnson, L Steven, Sean R Eddy, & Elon Portugaly (2010). “Hidden Markov model speed heuristic and iterative HMM search procedure”. In: *BMC bioinformatics* 11.1, p. 431.

Joshi, Kandarp Rakeshkumar, Heena Dhiman, & Vinod Scaria (2014). “tbvar: a comprehensive genome variation resource for Mycobacterium tuberculosis”. In: *Database* 2014, bat083.

Karakousis, Petros C et al. (2004). “Dormancy phenotype displayed by extracellular Mycobacterium tuberculosis within artificial granulomas in mice”. In: *The Journal of experimental medicine* 200.5, pp. 647–657.

Kinnings, Sarah L et al. (2010). “The Mycobacterium tuberculosis drugome and its polypharmacological implications”. In: *PLoS computational biology* 6.11, e1000976.

Koul, Anil et al. (2007). “Diarylquinolines target subunit c of mycobacterial ATP synthase”. In: *Nature chemical biology* 3.6, pp. 323–324.

Krasowski, Agata et al. (2011). “DrugPred: a structure-based approach to predict protein druggability developed using an extensive nonredundant data set”. In: *Journal of chemical information and modeling* 51.11, pp. 2829–2842.

Krick, Teresa et al. (2014). “Amino acid metabolism conflicts with protein diversity”. In: *Molecular biology and evolution* 31.11, pp. 2905–2912.

Le Guilloux, Vincent, Peter Schmidtke, & Pierre Tuffery (2009). “Fpocket: an open source platform for ligand pocket detection”. In: *BMC bioinformatics* 10.1, p. 168.

Li, Jianrong et al. (1997). “Nitric oxide reversibly inhibits seven members of the caspase family via S-nitrosylation”. In: *Biochemical and biophysical research communications* 240.2, pp. 419–424.

Li, Weizhong, & Adam Godzik (2006). “Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences”. In: *Bioinformatics* 22.13, pp. 1658–1659.

Lunardi, Juleane et al. (2013). “Targeting the Histidine Pathway in Mycobacterium tuberculosis”. In: *Current topics in medicinal chemistry* 13.22, pp. 2866–2884.

Ma, Qingjun et al. (2006). “The Mycobacterium tuberculosis LipB enzyme functions as a cysteine/lysine dyad acyltransferase”. In: *Proceedings of the National Academy of Sciences* 103.23, pp. 8662–8667.

Melo, Francisco, & Andrej Sali (2007). “Fold assessment for comparative protein structure modeling”. In: *Protein Science* 16.11, pp. 2412–2426.

Monera, Oscar D et al. (1995). “Relationship of sidechain hydrophobicity and α-helical propensity on the stability of the single-stranded amphipathic α-helix”. In: *Journal of peptide science* 1.5, pp. 319–329.

Movahedzadeh, Farahnaz et al. (2004). "The Mycobacterium tuberculosis ino1 gene is essential for growth and virulence". In: *Molecular microbiology* 51.4, pp. 1003–1014.

Murphy, Dennis J, & James R Brown (2007). "Identification of gene targets against dormant phase Mycobacterium tuberculosis infections." In: *BMC infectious diseases* 7, p. 84. ISSN: 1471-2334. DOI: 10.1186/1471-2334-7-84. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1950094%5C&tool=pmcentrez%5C&rendertype=abstract.

Muttucumaru, DG et al. (2004). "Gene expression profile of¡ i¿ Mycobacterium tuberculosis¡/i¿ in a non-replicating state". In: *Tuberculosis* 84.3, pp. 239–246.

Newton, Gerald L, & Robert C Fahey (2002). "Mycothiol biochemistry". In: *Archives of microbiology* 178.6, pp. 388–394.

Ohno, Hideaki et al. (2003). "The effects of reactive nitrogen intermediates on gene expression in Mycobacterium tuberculosis". In: *Cellular microbiology* 5.9, pp. 637–648.

Ouellet, Hugues et al. (2009). "Reaction of Mycobacterium tuberculosis Cytochrome P450 Enzymes with Nitric Oxide†". In: *Biochemistry* 48.5, pp. 863–872.

Perola, Emanuele, Lee Herman, & Jonathan Weiss (2012). "Development of a rule-based method for the assessment of protein druggability". In: *Journal of chemical information and modeling* 52.4, pp. 1027–1038.

Pérot, Stéphanie et al. (2010). "Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery". In: *Drug discovery today* 15.15, pp. 656–667.

Porter, Craig T, Gail J Bartlett, & Janet M Thornton (2004). "The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data". In: *Nucleic acids research* 32.suppl 1, pp. D129–D133.

Radi, Rafael (2004). "Nitric oxide, oxidants, and protein tyrosine nitration". In: *Proceedings of the National Academy of Sciences* 101.12, pp. 4003–4008.

Radusky, Leandro G et al. (2015). "An integrated structural proteomics approach along the druggable genome of Corynebacterium pseudotuberculosis species for putative druggable targets". In: *BMC Genomics* 16.Suppl 5, S9.

Radusky, Leandro et al. (2014). "TuberQ: a Mycobacterium tuberculosis protein druggability database". In: *Database* 2014, bau035.

Raman, Karthik, & Nagasuma Chandra (2008). “Mycobacterium tuberculosis interactome analysis unravels potential pathways to drug resistance”. In: *BMC microbiology* 8.1, p. 234.

Raman, Karthik, Preethi Rajagopalan, & Nagasuma Chandra (2005). “Flux balance analysis of mycolic acid pathway: targets for anti-tubercular drugs”. In: *PLoS computational biology* 1.5, e46.

Raman, Karthik, Kalidas Yeturu, & Nagasuma Chandra (2008). “targetTB: a target identification pipeline for Mycobacterium tuberculosis through an interactome, reactome and genome-scale structural analysis”. In: *BMC systems biology* 2.1, p. 109.

Rao, S Trivikrama, & Michael G Rossmann (1973). “Comparison of super-secondary structures in proteins”. In: *Journal of molecular biology* 76.2, pp. 241–256.

Rengarajan, Jyothi, Barry R Bloom, & Eric J Rubin (2005). “Genome-wide requirements for Mycobacterium tuberculosis adaptation and survival in macrophages”. In: *Proceedings of the National Academy of Sciences of the United States of America* 102.23, pp. 8327–8332.

Robinson, Jonathan L, Kristin J Adolfsen, & Mark P Brynildsen (2014). “Deciphering nitric oxide stress in bacteria with quantitative modeling”. In: *Current opinion in microbiology* 19, pp. 16–24.

Sassetti, Christopher M, Dana H Boyd, & Eric J Rubin (2003). “Genes required for mycobacterial growth defined by high density mutagenesis”. In: *Molecular microbiology* 48.1, pp. 77–84.

Sassetti, Christopher M, & Eric J Rubin (2003). “Genetic requirements for mycobacterial survival during infection”. In: *Proceedings of the National Academy of Sciences* 100.22, pp. 12989–12994.

Schmidtke, Peter, & Xavier Barril (2010). “Understanding and predicting druggability. A high-throughput method for detection of drug binding sites”. In: *Journal of medicinal chemistry* 53.15, pp. 5858–5867.

Schnappinger, Dirk et al. (2003). “Transcriptional adaptation of Mycobacterium tuberculosis within macrophages insights into the phagosomal environment”. In: *The Journal of experimental medicine* 198.5, pp. 693–704.

Sheridan, Robert P et al. (2010). "Drug-like density: a method of quantifying the "bindability" of a protein target based on a very large set of pockets and drug-like ligands from the Protein Data Bank". In: *Journal of chemical information and modeling* 50.11, pp. 2029–2040.

Singhal, Anshika et al. (2013). "Regulation of homocysteine metabolism by Mycobacterium tuberculosis S-adenosylhomocysteine hydrolase". In: *Scientific reports* 3.

Spalding, Maroya D, & Sean T Prigge (2010). "Lipoic acid metabolism in microbial pathogens". In: *Microbiology and Molecular Biology Reviews* 74.2, pp. 200–228.

Talaat, Adel M et al. (2004). "The temporal expression profile of Mycobacterium tuberculosis infection in mice". In: *Proceedings of the National Academy of Sciences of the United States of America* 101.13, pp. 4602–4607.

Volkamer, Andrea, Axel Griewel, et al. (2010). "Analyzing the topology of active sites: on the prediction of pockets and subpockets". In: *Journal of chemical information and modeling* 50.11, pp. 2041–2052.

Volkamer, Andrea, Daniel Kuhn, Thomas Grombacher, et al. (2012). "Combining global and local measures for structure-based druggability predictions". In: *Journal of chemical information and modeling* 52.2, pp. 360–372.

Volkamer, Andrea, Daniel Kuhn, Friedrich Rippmann, et al. (2012). "DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment". In: *Bioinformatics* 28.15, pp. 2074–2075.

Voskuil, Martin I, Iona L Bartek, et al. (2011). "The response of Mycobacterium tuberculosis to reactive oxygen and nitrogen species". In: *Frontiers in microbiology* 2.

Voskuil, Martin I, Dirk Schnappinger, et al. (2003). "Inhibition of respiration by nitric oxide induces a Mycobacterium tuberculosis dormancy program". In: *The Journal of experimental medicine* 198.5, pp. 705–713.

Wehenkel, Annemarie et al. (2006). "The structure of PknB in complex with mitoxantrone, an ATP-competitive inhibitor, suggests a mode of protein kinase regulation in mycobacteria". In: *FEBS letters* 580.13, pp. 3018–3022.

Winzeler, Elizabeth A et al. (1999). "Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis". In: *science* 285.5429, pp. 901–906.

Zambrowicz, Brian P, & Arthur T Sands (2004). “Modeling drug action in the mouse with knockouts and RNA interference”. In: *Drug Discovery Today: TARGETS* 3.5, pp. 198–207.

Capítulo 4

Comparación estructural y mecanística de la familia de Cyclopropane Mycolic Acid Synthases (CMAS): un compuesto esencial de la pared de *Mtb.*

4.1 Introducción

Las CMAS, son las responsables de producir las modificaciones a los dobles enlaces de los ácidos micólicos (AM) inmaduros . Son enzimas con actividad metiltransferasa cuyas modificaciones se producen al transferir un grupo metilo del donor S-adenosil-L-metionina (SAM) al ligando en cuestión, un AM inmaduro. Todas presentan el típico motivo de unión de a nucleótidos conocido como *Rossman Fold*, que une SAM y otro sub-dominio encargado -de unir al lípido en cuestión que es típico de las metiltransferasas de moléculas orgánicas pequeñas y lípidos (Defelipe et al., 2011; Martin, & McMillan, 2002). Ver Figura **??**

Debido a la dificultad de trabajar con estas proteínas *in vitro*, principalmente desde el punto de vista bioquímico, la mayor parte de las determinaciones de actividad de estas enzimas fueron corroboradas mediante experimentos mutacionales (*knock-out*) en *Mtb*, crecimiento, con una posterior extracción y análisis de los AM resultantes por cromatografía en capa delgada y/o Resonancia Magnética Nuclear. (Yuan, Lee, et al., 1995; Yuan, & Barry, 1996; Yuan, Crane, et al., 1997; Behr et al., 2000; Glickman, Cahill, & Jacobs, 2001; Glickman, 2003; Barkan

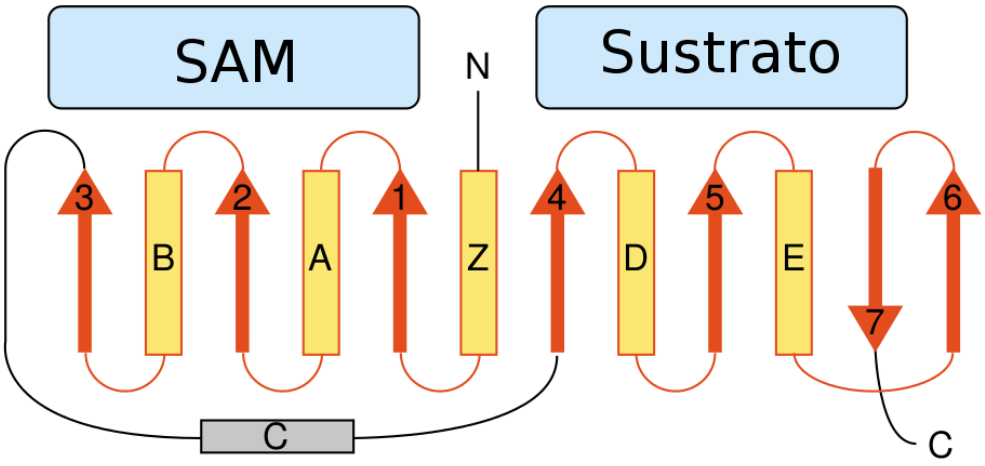


Figura 4.1: **Representación de guardas del plegado núcleo de las metiltransferasas dependientes de SAM.** Adaptado de (Martin, & McMillan, 2002)

et al., 2010).

El mecanismo de reacción propuesto (Marrakchi, Lanéelle, & Daffé, 2014; Liao et al., 2011) se puede describir en dos etapas: La primera consiste en la transferencia del grupo metilo del SAM al doble enlace (mediante un mecanismo asociativo) formando un carbocatión secundario. La segunda, dependiente de cada enzima, consiste en la resolución del carbocatión hacia: i) un ciclopropano (en cis en el caso de cmaA1, pcaA y mmaA2 mientras que cmaA2 es promiscua y puede producir tanto cis como trans), ii) un doble enlace en *trans* con un metilo vecinal (mmaA1) o iii) un metil-alcohol (mmaA4) **??**. MmaA3 resulta en un caso particular ya que su sustrato propuesto es el hidroxi micolico producido por mmaA4 que luego es metilado por mmaA3, produciendo un metil-eter. Un resumen de la actividad descripta de cada enzima sobre el ácido micólico inmaduro se puede observar en la figura **??**.

El objetivo de este capitulo es comprender, mediante el uso de herramientas bioinformáticas y de simulación computacional, los determinantes moleculares de la actividad diferencial de las distintas CMAS que les permite realizar las diversas modificaciones químicas a los grupos olefina y alcohol (en el caso de mmaA3) del AM. Para realizar estas tareas modelaremos las estructuras de umaA, ufaA1, mmaA1, mmaA3 y Rv3720 y estudiaremos mediante técnicas de dinámica molecular y QM/MM el mecanismo de reacción de cmaA2 y mmaA4. Por otra parte analizaremos el mecanismo de las CMAS en general mediante comparación estructural y funcional.

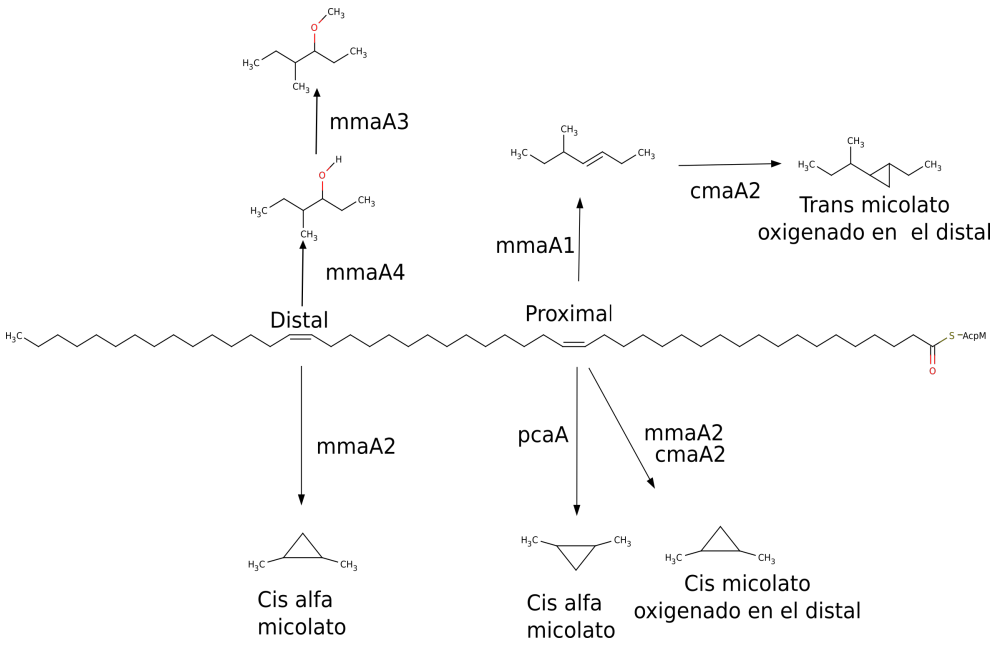


Figura 4.2:
Actividades propuestas para cada una de las CMAS en base a resultados experimentales. (Yuan, & Barry, 1996; Behr et al., 2000; Glickman, Cahill, & Jacobs, 2001; Glickman, 2003; Barkan et al., 2010)

4.2 Materiales y métodos

4.2.1 Alineamiento múltiple de secuencias

El alineamiento múltiple de secuencias es aún un problema abierto en la bioinformática. En general se utilizan heurísticas basadas en programación dinámica para generar alineamientos de buena calidad. Dado que se trata de un problema complejo de resolver se decidió utilizar una estrategia mixta de alineamiento múltiple que se encuentra implementada en T-COFFEE. (Poirot, O'Toole, & Notredame, 2003) . Sucintamente, el algoritmo realiza todos los alineamientos de a pares posibles y utiliza la información obtenida de ellos como ayuda para poder producir alineamientos múltiples de mejor calidad. En este trabajo se utilizó la versión implementada como servicio web `http://www.tcoffee.org/`.

4.2.2 Modelado comparativo

El modelado comparativo (o por homología) se realizó de forma similar al utilizado para el desarrollo de TuberQ (Radusky et al., 2014) pero con un control manual lo que permite una corrección de los alineamientos si hubiese problemas. Para umaA (Q6MX39) se utilizó como molde el cristal 1L1E (pcaA de *Mtb*) con 59% de identidad, 73% de positivos y una cobertura

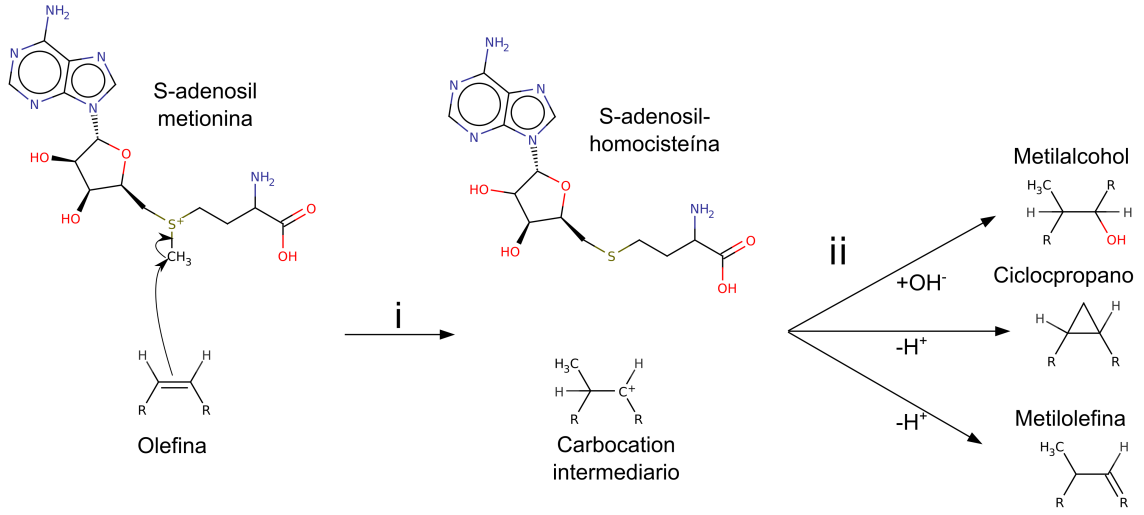


Figura 4.3: **Mecanismo de reacción general propuesto para las CMAS de *Mycobacterium tuberculosis*.** La reacción se divide en dos pasos elementales: i) El ataque y formación del metil-carbocatión, común a todas las CMAS y ii) la resolución del mismo a distintos productos, propia de cada CMAS.

del 99%. Se utilizó como molde el cristal 1KPG (cmaA1 de *Mtb*) para las siguientes proteínas:

- ufaA (O53732, 33% de identidad, 46% de positivos y 63% de cobertura)
- mmaA1 (P9WPB1, 56% de identidad, 69% de positivos y 99% de cobertura)
- mmaA3 (P0CH91, 66% de identidad, 79% de positivos y 94% de cobertura)

Por último para *Probable fatty acid methyltransferase Rv3720* (O69687) se utilizó como molde el cristal 1TPY (mmaA2 de *Mtb*) con un 32% de identidad, 46% de positivos y una cobertura del 57%.

4.2.3 Alineamiento estructural

Teniendo en cuenta que proteínas con una identidad de secuencia baja pueden presentar el mismo plegado y por ende se espera que haya residuos importantes para el plegado que se encuentran más conservados. Uno de los objetivos es poder analizar una familia de proteínas (CMAS de *Mycobacterium tuberculosis*) es necesario contar con un método de alineamiento estructural que sea independiente de la identidad de secuencia. En este capítulo e l alineamiento estructural se realizó con MAMMOTH que utiliza una heurística dividida en 4 cuatro pasos: i) Generar alineamientos estructurales de carbonos alfa de heptapéptidos de las estructuras de interés (todos contra todos). ii) Utilizando programación dinámica generar un alineamiento que

maximice el puntaje de similitud. iii) Encontrar el subconjunto de alineamientos que tienen sus correspondientes carbonos alfa cercanos en el espacio cartesiano y iv) Evaluar mediante un cálculo de p-valor la probabilidad de realizar los alineamientos al azar.

4.2.4 Estimación del perfil la energía libre de reacción de cmaA2 y mmaA4

Armado del sistema y parametrización de ligandos

Tanto la S-adenosil-L-metionina,la S-adenosil-L-homocisteina como los lípidos modelo utilizados no se encuentran parametrizadas en AMBER por lo cual fue necesario computar las cargas RESP (con Gaussian 03, HF/6-31G*) a partir de una geometría optimizada de los compuestos. Los parámetros de unión fueron tomados del GAFF (Wang et al., 2004). Mediante *docking* el lípido y el SAM fueron posicionados en el sitio activo de cmaA2 (PDBID:1KPI) o de mmaA4 (PDBID:3HA5) utilizando como guía los ligandos presentes en los respectivos cristales. Los detalles de los parámetros de pueden observar en el Anexo B. Se protonó el sistema con tLeap (Case et al., 2014) y agrego una caja de aguas TIP3P 10 Åde radio.

Dinámica molecular clásica de cmaA2 y mmaA4

Para el sistema reactivo inicial (cmaA2+SAM+hidrocarburo+HCO3) se procedió a minimizar los sistemas con SANDER (Case et al., 2014) durante 10000 pasos, en condiciones periódicas de borde a volumen constante y con un corte de las interacciones de largo alcance a 10 Å. Luego se procedió a un calentado suave del sistema desde 10K a 300K durante 100ps con el termostato de Berendsen (constante de tiempo 2 ps). Por último se procedió a equilibrar el sistema a 300K y 1 bar (Barostato de Berendsen, constante de tiempo 2ps) realizando simulaciones en el ensamble NPT durante 100ps. Se simularon 150ns de dinámica molecular para tomar fotos que sean aptas como punto de partida para realizar MSMD. Se realizó un procedimiento similar para el sistema intermediario (cmaA2+SAH+carbocation+HCO3) pero solo corriendo 10ns para obtener fotos como punto de partida. Para el caso de mmaA4 el protocolo utilizado fue similar, una vez equilibrado el sistema se corrieron 500ns de dinámica molecular clásica y se tomaron fotos para realizar MSMD.

Dinámica molecular híbrida de cmaA2 y mmaA4

En el caso de cmaA2 el sistema reactivo inicial se definió el sistema cuántico como los átomos de azufre, y los tres carbonos adyacentes al mismo con sus hidrógenos en el caso del SAM mientras que el hidrocarburo se eligieron los átomos que forman el doble enlace del mismo y todos los carbonos e hidrógenos que estén a 5 enlaces del mismo, en total el sistema cuántico cuenta con 33 átomos más 4 *link atoms*, un total de 37 átomos para ser tratados de forma cuántica. El hamiltoniano empleado fue DFTB, que brinda un balance entre la calidad del cálculo y la velocidad de computo. La estrategia para llevar a régimen al sistema es similar a la utiliza en la dinámica molecular clásica. Minimizar (2000 pasos, en condiciones periódicas de borde), calentar (de 10K a 300K, Berendsen, constante de acoplamiento 1 ps, 50ps de simulación), equilibrar (300K. Termostato de Langevin, 50ps de simulación). A partir de ahí se realiza una simulación de 500ps con la coordenada de reacción fija (ver ecuaciones) guardando las fotos cada 12.5 ps siendo estos los puntos de partida.

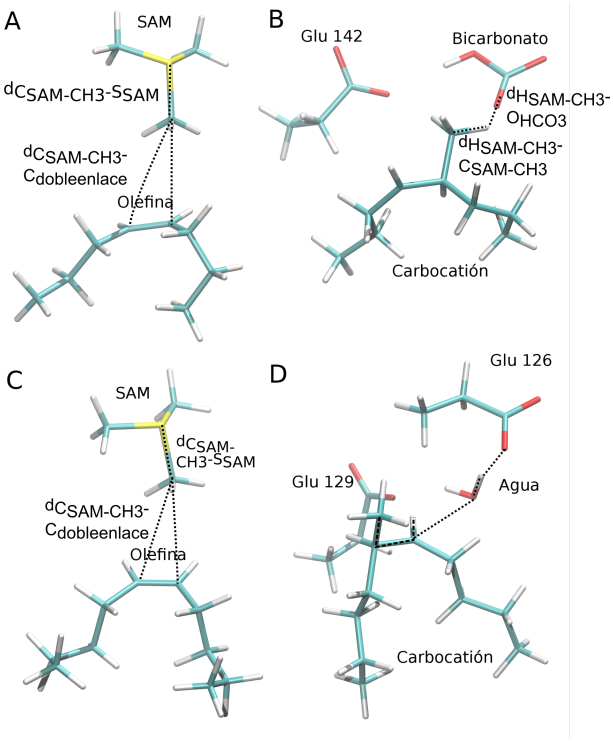


Figura 4.4: **Sistemas QM utilizado para cada una de las simulaciones:** A) Primer paso de la reacción en cmaA2. Se colocaron 4 *link atoms* para modelar la interfase entre los sistemas QM y MM, dos en los carbonos del SAM y dos en los carbonos de la olefina. B) Segundo paso de la reacción en cmaA2. Se colocaron 3 *link atoms*, dos en el carbocatión (en la misma posición que el paso anterior) y uno en el glutámico. C) Primer paso de la reacción en mmaA4. Se colocaron 2 *link atoms* en los carbonos del SAM. D) Segundo paso de la reacción en mmaA4. Se colocaron 2 l*ink atoms* en los glutámicos 126 y 129. Las lineas punteadas indican distancias utilizadas en las coordenadas de reacción, ver texto

Para el primer paso de la reacción se utilizó un tiempo de integración de 1 fs. y una estrategia de relajación del sistema clásico (Hybrid Differential Relaxation Algorithm, HyDRA (Ramirez et al., 2014)) desarrollada en el grupo. Este algoritmo permite relajar al sistema clásico mientras se mantiene rígido al sistema cuántico, evitando el sobrecalentamiento que ocurre si se mueva la coordenada de reacción de forma muy veloz. Se realizaron 50000 pasos de totales de simulación (12500 de QM-MM). La coordenada de reacción (ecuación **??**) utilizada fue la siguiente:

$$Coordenada = d(C_{SAM-CH3} - S_{SAM}) - d(C_{SAM-CH3} - C_{DobleEnlace}) \tag{4.1}$$

donde $C_{SAM-CH3}$ es el metilo a ser transferido, S_{SAM} es el átomo de azufre del SAM y $C_{DobleEnlace}$ es alguno de los átomos de carbono que forman el doble enlace. La velocidad guía fue de 0.04 Åpor ps. Ver figura **??**

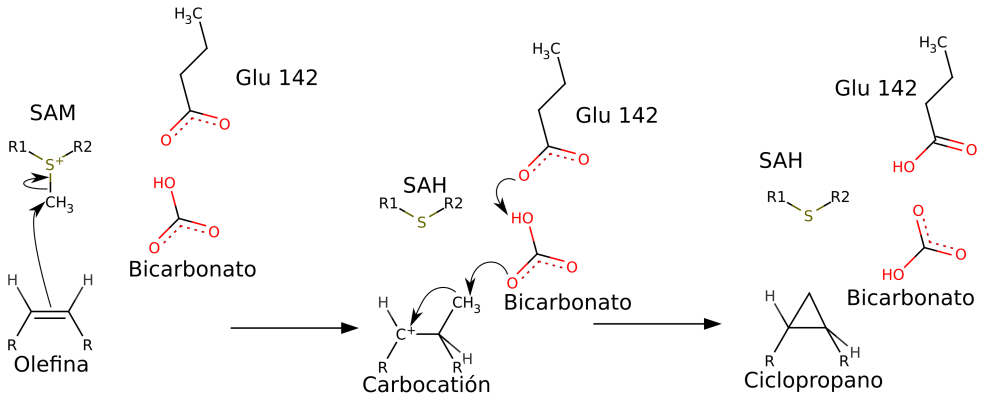


Figura 4.5: **Mecanismo de reacción propuesto para cmaA2.** Las flechas indican el movimiento de los electrones.

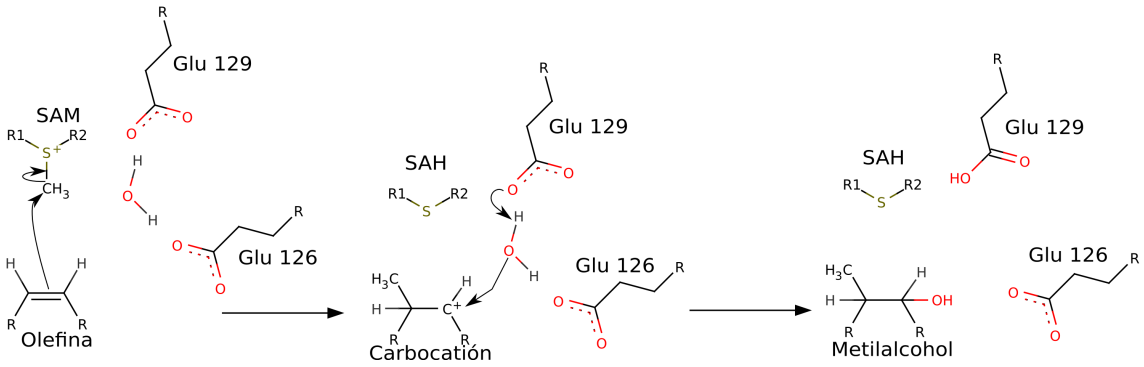


Figura 4.6: **Mecanismo de reacción propuesto para mmaA4.** Las flechas indican el flujo de los electrones.

Para el segundo paso, el sistema cuántico elegido consistió HCO_3, la cadena lateral del Glu 137 y los mismos carbonos del ahora carbocatión. El sistema consiste en 40 átomos cuánticos +

3 *link atoms*. La reacción se realiza sin HyDRA con un paso de integración de 0,5 fs. y durante 50000 pasos. La coordenada de reacción fue la siguiente (Ver Figura **??** para cmaA2 y **??** para mmaA4, los átomos involucrados en la coordenada se encuentra mostrados en la figura **??**):

$$Coordenada = d(H_{SAM-CH3} - C_{SAM-CH3}) - d(H_{SAM-CH3} - O_{HCO3}) \tag{4.2}$$

Donde $H_{SAM-CH3}$ es el protón a ser transferido y O_{HCO3} el oxígeno receptor. La velocidad guía fue de 0.08 Åpor ps.

En el caso de mmaA4 el sistema cuántico estuvo compuesto de los átomos de azufre, y los tres carbonos adyacentes al mismo con sus hidrógenos en el caso del SAM y la olefina completa (dado que se trata de una más corta que en el caso de cmaA2. El protocolo de simulación QM-MM MSMD es idéntico (tiempo de simulación, paso de integración, hamiltoniano, coordenada de reacción), incluso la cantidad de pasos de relajación en HyDRA. En el caso del segundo paso el sistema cuántico se encuentra conformado por el ligando carbocationico con su metilo representado en forma cuántica, la cadena lateral de un glutámico (el 126, equivalente a la posición del bicarbonato de cmaA2 y otras CMAS) y una molécula de agua.

4.3 Resultados

Comenzaremos esta sección comentando los resultados bioinformáticos (modelado por homología, alineamiento de secuencia y alineamientos estructurales) que comprende el punto de partida para el estudio de posibles mecanismos de reacción. Luego realizamos la estimación de los perfiles de energía libre de cmaA2 y de mmaA4 para comprender sus mecanismos de reacción posibles y comprender las diferencias en selectividad y reactividad observadas *in vivo* a la luz de los resultados presentados en este capítulo. En este capítulo utilizaremos la nomenclatura de la estructura secundaria utilizada por Huang y colaboradores (Huang et al., 2002) derivada de la nomenclatura general propuesta para las metiltransferasas (Martin, & McMillan, 2002). En la figura **??** se muestra dicha nomenclatura.

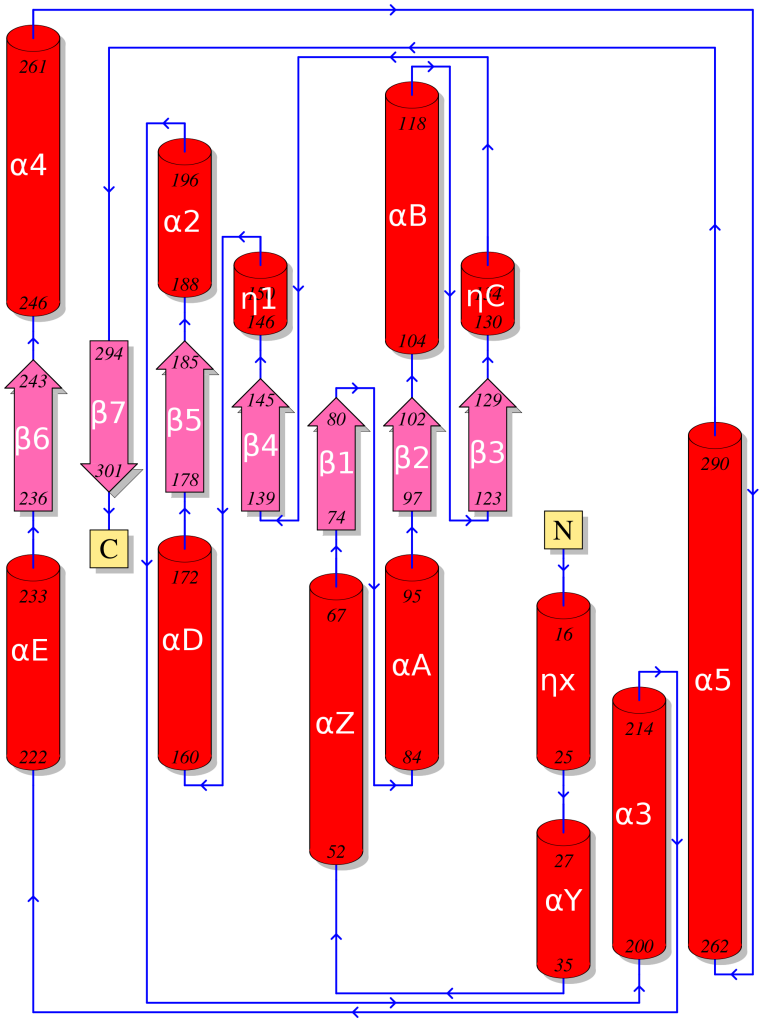


Figura 4.7: **Nomenclatura de la estructura secundaria de las *Cyclopropane Mycolic Acid Synthses***. Basado en (Huang et al., 2002)

4.3.1 Modelado comparativo, estructura del sitio activo y de unión

Como primer paso se decidió realizar un modelado comparativo de las CMAS de las cuales no se disponen estructura(umaA, ufaA1, mmaA1, mmaA3 y rv3720). Para ello se utilizó un protocolo similar al realizado para el modelado a escala genómica de todo el proteoma de *Mtb* que ya fue explicado en el capítulo de Métodos. En la figura **??** se puede ver los resultados de cada modelo comparado con su molde y en la figura **??** los alineamientos de a pares entre el molde y la secuencia modelada.

En general se observan pocas diferencias en la posición de la cadena principal de los modelos generados respecto al molde. En el caso de umaA las diferencias se centran en la presencia de la hélice ηX, desestructurada en pcaA con una parte no resuelta y en la estructuración y aparición de la hélice $\eta 1$. Para el caso de ufaA, mmaA1 y mmaA3 pocas diferencias relevantes se observan en los modelos, la única destacable es la posición del loop entre las hélices $\alpha 2$ y $\alpha 3$

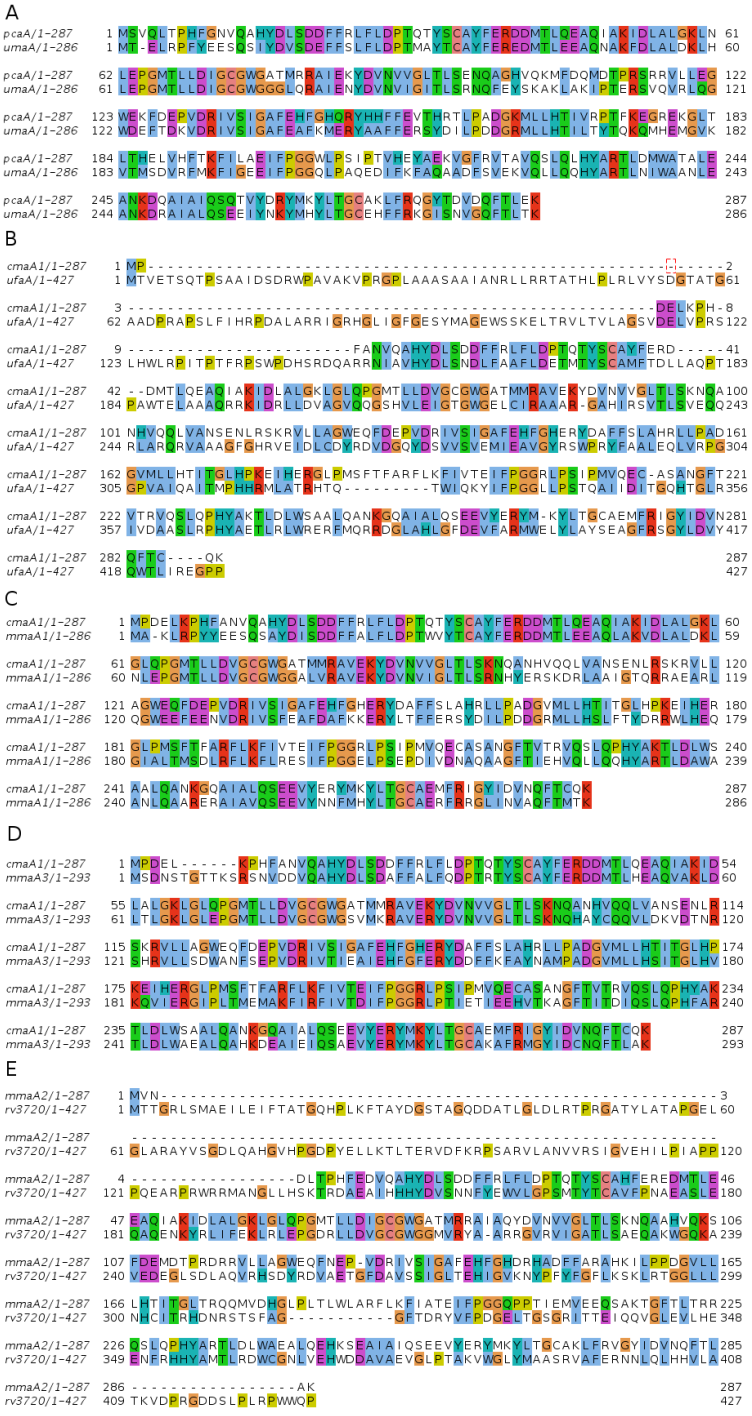


Figura 4.8: **Alineamiento entre el molde y su secuencia objetivo para el modelado de: A)umaA, B) ufaA, C) mmaA1, D) mmaA3 y E) rv3720.**

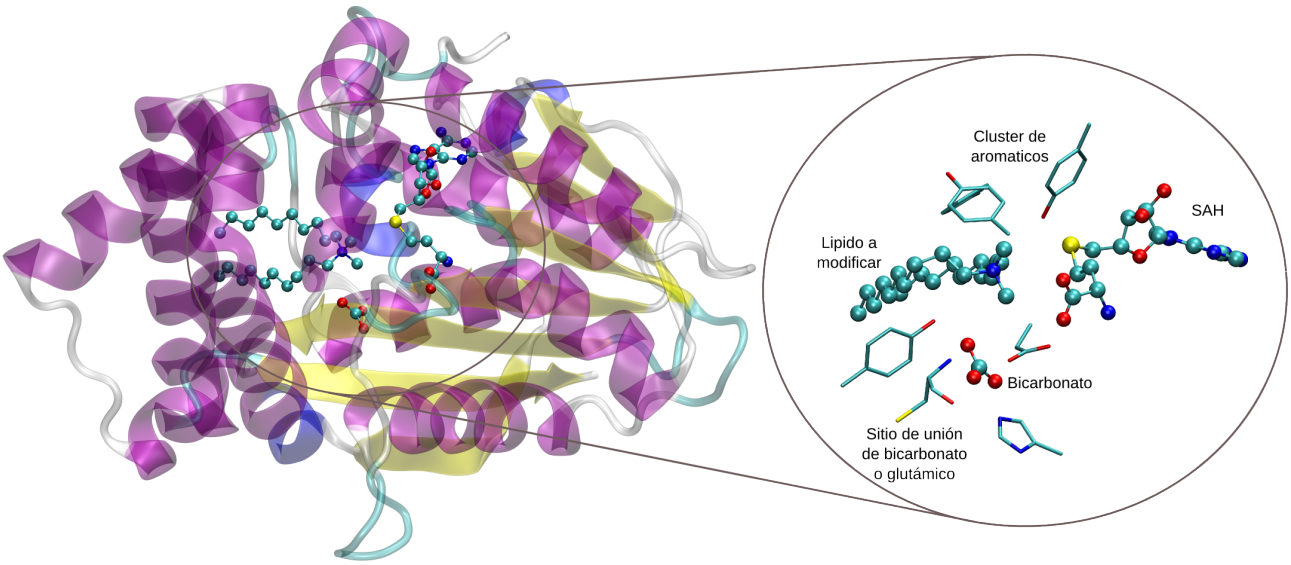


Figura 4.9: **Estructura general de las CMAS (utilizando a cmaA2 como modelo - PDBID:1KPI) con una ampliación mostrando los componentes de su sitio activo.**

en el caso de mmaA1 y mmaA3. UfaA no presenta la formación de la hélice $\alpha 2$. En rv3720 se observan las diferencias más significativas, como era de esperar por la cobertura y la identidad entre el modelo y la secuencia a modelar(57% y 32 % respectivamente), la más interesante se trata de la posición de la hélice ηX que se tuerce. Otra de las diferencias se encuentra en el sitio activo dado que el modelo generado no presenta la hélice ηC donde se posiciona la histidina responsable de coordinar al bicarbonato en los otros miembros de la familia.

Como se puede apreciar en las figuras **??** y **??**, el plegado típico de las CMAS esta conformado por 7 hojas beta (6 de ellas paralelas y una antiparalela) y 13 hélices alfa de largo variable. El sub-dominio de unión a nucleótido característico de las metiltransferasas esta formado principalmente por las hojas beta mientras que el dominio de unión a los lípidos esta conformado por las hélices alfa. El sitio activo se encuentra conformado por un *cluster* de residuos aromáticos (tirosinas y fenilalaninas), el sitio de unión a SAM/SAH, conformado por residuos polares que interaccionan con la adenenina y la ribosa para estabilizarla y anclarla como así también residuos de la cadena principal en el caso de la parte aminoacídica. Por último, la característica primordial de este grupo de enzimas es la presencia de un sitio de unión a bicarbonato/carbonato conformado por una histidina, el nitrógeno de la cadena principal de una cisteína y una tirosina que se muestra en la figura **??**.

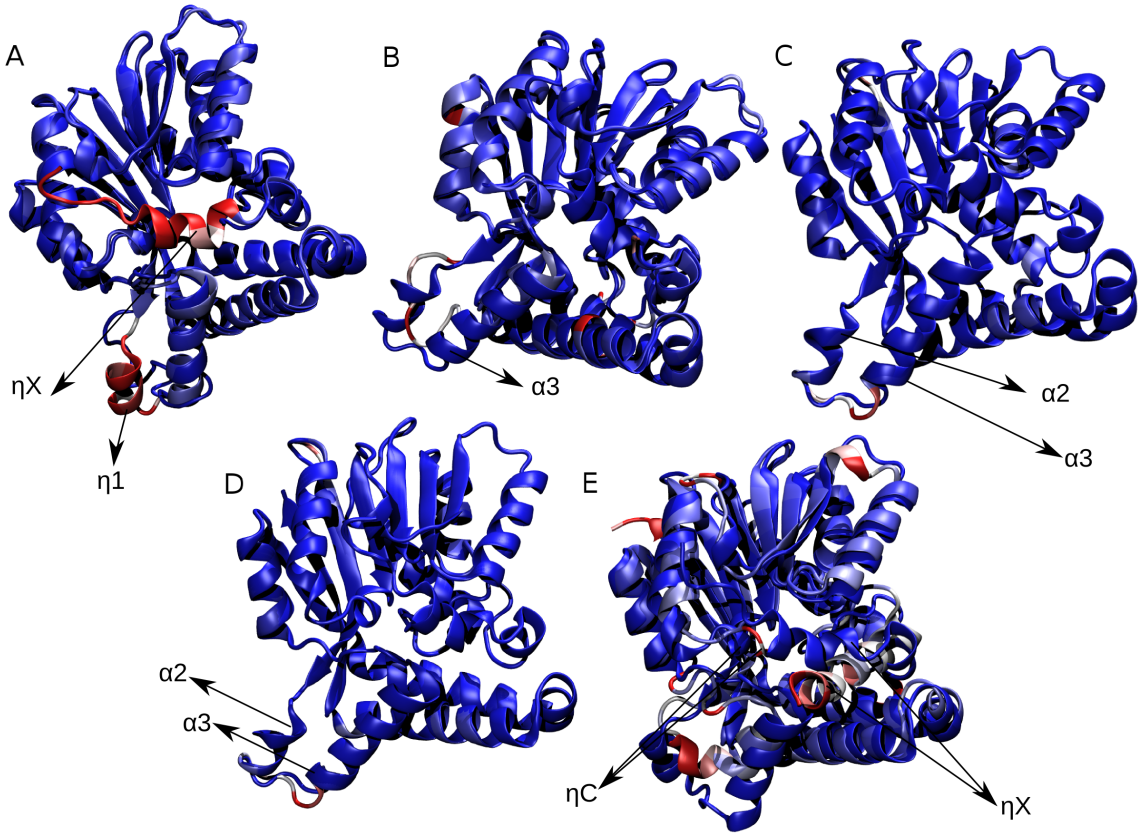


Figura 4.10: **Comparación estructural entre los modelos generados de las CMAS y los moldes utilizados:** A) pcaA-umaA, B) cmaA1-ufaA, C) cmaA1-mmaA1, D) cmaA1-mmaA3 y E) mmaA2 y rv3720. El código de color indica diferencias en el RMSD de la cadena principal siendo la escala de azul (bajo RMSD) a rojo (alto RMSD).

4.3.2 Comparación entre las distintas CMAS

Se alinearon las 9 secuencias de las CMAS para encontrar aminoácidos conservados, principalmente en el sitio activo y de unión de los sustratos. En primer lugar, realizamos un alineamiento de secuencia de todas las CMAS presentes en el genoma de *M. tuberculosis*. Esta determinación se realizó a partir de la familia de PFAM de las mismas (PFAMId: 02353). El resultado se encuentra en la figura **??** con un árbol de agrupamiento por identidad de secuencia presentado en la figura **??**.

Al agrupar las secuencias por su similitud global utilizando la matriz de sustitución BLOSUM62 se pueden observar varios fenómenos interesantes. El primero es la formación de un conjunto externo a las CMAS más descriptas conformado por ufaA1 y rv3720, esto es de esperar ya que se tratan de las dos proteínas con un largo muy distinto debido a la presencia de 130-140 aminoácidos más en la región N-terminal. La función de esa inserción es desconocida y su estructura no pudo ser modelada por carencia de molde.

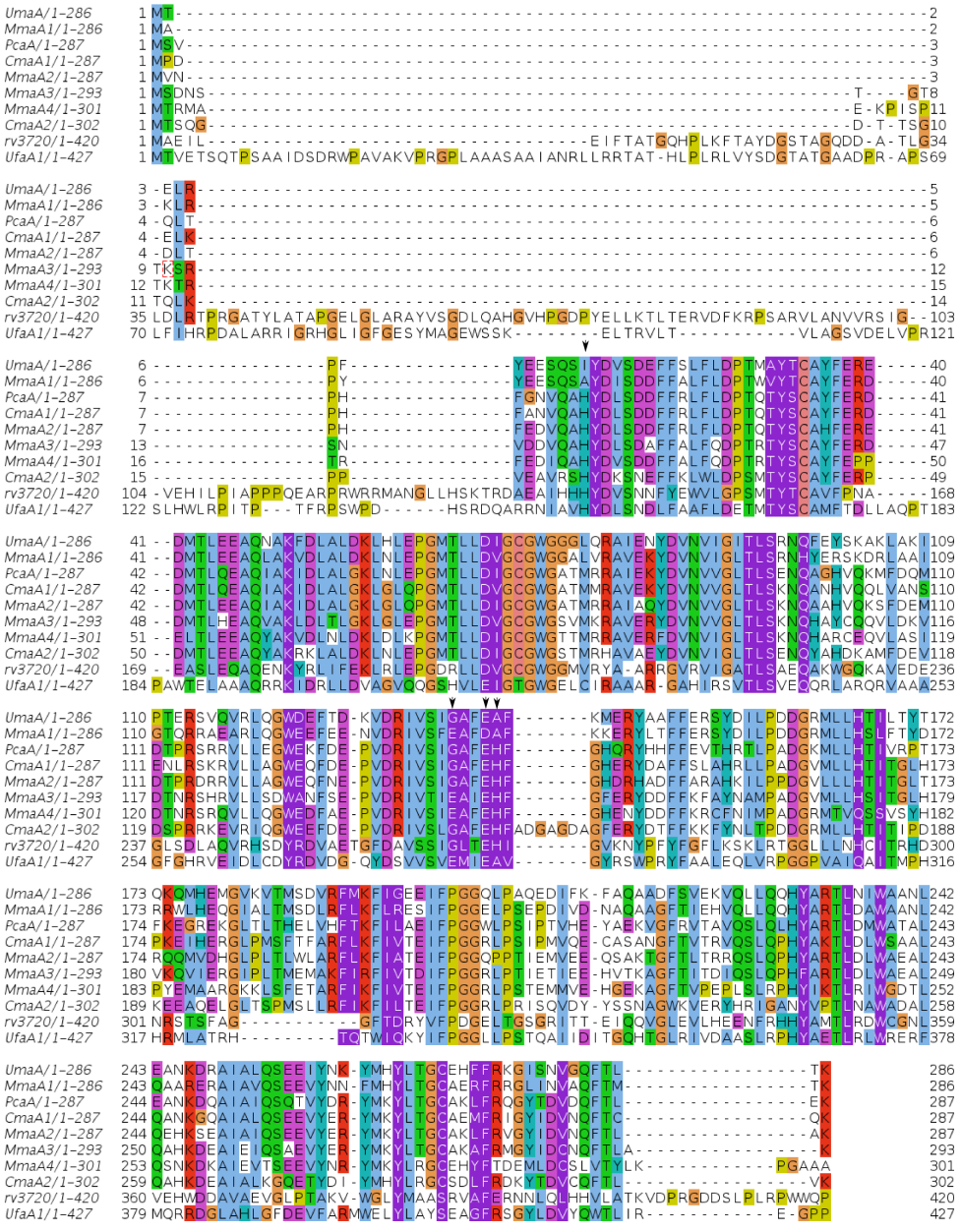


Figura 4.11: Alineamiento de secuencia de las proteínas con dominio CMAS (según PFAM) en el genoma de *M. tuberculosis*. Los colores corresponden a los utilizados por ClustalX. Los residuos involucrados en unión o que son parte del sitio activo se encuentran destacados en violeta. Las flechas indican posiciones importantes destacadas en el texto.

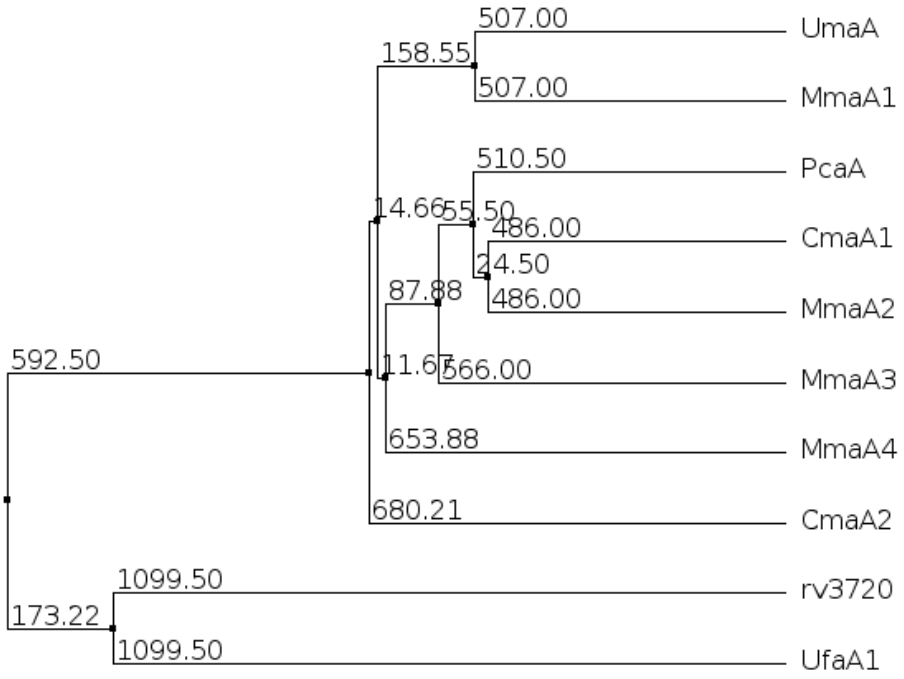


Figura 4.12: Arbol guía producido por UPGMA a partir del alineamiento utilizando como puntuación la matriz de sustitución BLOSUM62.

Otro resultado interesante de esta comparación resulta de la aparición de otro agrupamiento formado por umaA y mmaA1, dado que se conoce la actividad de mmaA1 pero desconoce la de umaA puede resultar de interés este agrupamiento, lo analizaremos en más detalle en las secciones que siguen. Existen otro grupo conformado por las CMAS canónicas pcaA, cmaA1 y mmaA2 mientras que cmaA2 permanece fuera de este grupo. Entre estos grupos se ubican mmaA3 y mmaA4, encargadas de metilar e introducir la porción oxigenada en el doble enlace distal. (ver figura **??**)

Cuando en vez de comparar la identidad de secuencia global observamos los residuos importantes para la reactividad y la unión, la situación es otra. Por ejemplo, las CMAS canónicas (pcaA,cmaA1-2) presentan un sitio de unión a carbonato/bicarbonato descripto mediante cristalografía de rayos-X en estas proteínas que también se encuentra en el homologo más estudiado de *E. coli*. (Iwig, Uchida, et al., 2005). En estas enzimas existe un espacio formado por un residuo de glicina, la cadena principal de una cisteína y las cadenas laterales de una histidina y una tirosina. Al realizar un alineamiento estructural entre las estructuras disponibles en esa misma posición en mmaA1,3-4 y en ufaA1 se encuentra un residuo de ácido glutámico. (Ver figura **??** marcado con flechas y Figura **??**A para verlo sobre la estructura). Dado que el

glutámato es una peor base que el bicarbonato es posible que esta diferencia explique porque estas enzimas son incapaces de producir un ciclopropano y resuelven el carbocatión de otras maneras. Otra diferencia destacable que presenta mmaA1 con el restos de las CMAS, es la presencia de un residuo de aspártico en reemplazo de un glutámico presente en el resto de la familia en el sitio adyacente al bicarbonato (Ver figure **??** marcado con flechas y Figura **??**A para verlo sobre la estructura).

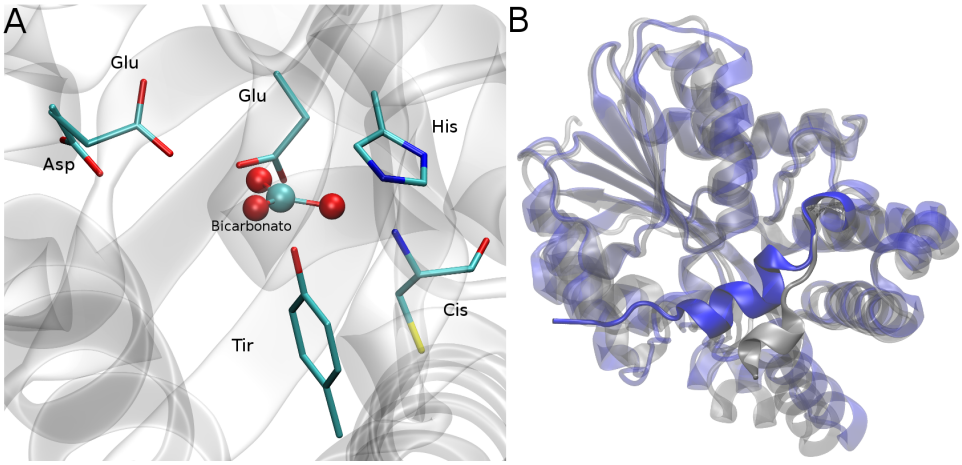


Figura 4.13: **Estructura de los sitios activos de las CMAS y del N-terminal de mmaA3 y mmaA4.** A) Estructura del sitio activo de cmaA2, mmaA1 y mmA4. En esferas y palitos se muestra la posición del ión bicarbonato en cmaA2. El resto de las posiciones corresponden a mmaA4, salvo el aspártico que pertenece a mmaA1. B) Estructura del N-terminal de mmaA4 (en gris) y mmaA3 (en azul).

Todas estas enzimas presentan en el sitio de unión un gran agrupamiento de aminoácidos aromáticos situados principalmente cerca del sitio activo (donde se posiciona la olefina para ser atacada por el SAM). En el sitio de unión a la olefina, en la hélice ηX se encuentra otra diferencia. Mientras que la mayor parte de las CMAS presentan una histidina en esa posición, mmaA1 presenta una alanina mientras que umaA presenta una isoleucina. Se desconoce que rol puede estar jugando este cambio en la especificidad del sustrato.

Desde el punto de vista estructural, cabe señalar que la hélice ηX en el caso de mmaA4 presenta una gran movilidad y no se encuentra resuelta en ausencia ni presencia de sustratos, algo que sucede en el caso de cmaA2 solamente cuando este se encuentra en presencia de SAM/SAH. Esta diferencia en movilidad/estructura puede explicar la entrada de aguas al sitio activo de mmaA4 con mayor facilidad respecto a mmaA3. Por otra parte la falta de estructuración del N-terminal en mmaA4 deviene en el posicionamiento diferencial de un residuo histidina y otro tirosina de la hélice ηX respecto a las otras CMAS, dado que en mmaA4 no

tiene estructura, se encuentra expuesto a solvente mientras que en el resto de las CMAS se encuentra mirando al sitio de unión. En mmaA3 la tirosina se encuentra mirando al interior de la cavidad, posiblemente facilitando la estabilización del alcohol mediante interacciones de puente hidrógeno. Al encontrar mayor dificultad para la entrada de aguas al sitio activo en mmaA3 el par de glutámicos señalado arriba (Ver figura **??**A) probablemente tenga un pKa muy corrido y esté favorecida la forma protonada facilitando la transferencia del grupo metilo en forma concertada con la protonación.

Por último, de las CMAS que no presentan un dominio adicional con función desconocida se encuentran cmaA1 y umaA. El sitio activo de cmaA1 resulta idéntico al sitio activo de cmaA2, con lo cual cabría hipotetizar que ambas enzimas pueden realizar el mismo tipo de reacciones. Sin embargo, la sobreexpresión de cmaA1 produce cambios en el perfil de ácidos micolicos en *Mtb* (Yuan, Lee, et al., 1995) pero no su ausencia (Glickman, 2003), con lo cual esta demostrado que tiene una actividad de cis-ciclopropano sintasa pero no su relevancia *in vivo*. Si se intenta realizar una predicción de la actividad de umaA respecto a otros miembros de la familia de las CMAS se pueden observar dos diferencias en el sitio de unión de los sustratos, dos sustituciones de histidina por alanina, una en la hélice ηX y otra en el *loop* entre la hoja $\beta 4$ y la hélice $\eta 1$. Dado que presenta el patrón de unión a bicarbonato intacto como se puede observar en la Figura **??** y en la inspección visual del modelo, cabe suponer que realiza algún tipo de actividad del estilo ciclopropano. Sin embargo, tanto cmaA1 como umaA1 no presentan cambio alguno en sus perfiles de ácidos micólicos en cepas con mutantes simples de estos genes. La falta de efecto de las mutantes *in vivo* puede deberse a efectos compensatorios debido a redundancia entre estas enzimas. Por ejemplo, en ausencia de mmaA2, cmaA2 puede reemplazarla en la producción de cis-ciclopropanos oxigenados aunque con una producción mucho menor (Barkan et al., 2010) pudiendo ser las relaciones más compensatorias mucho más complejas (doble redundancia, interacción con complejos proteícos de síntesis como FAS II, etc) que no pueden ser estudiadas mediante las técnicas computacionales disponibles hoy en día.

4.3.3 Mecanismos de reacción

Mecanismo de reacción de cmaA2

Estudiaremos el mecanismo de reacción de cmaA2, descripta como una transferencia de metilo y una ciclopropinación. El mecanismo podría ser similar al descripto para cmaA1 dado que *in vivo* se producen cambios similares en la estructura de los micólicos al sobreexpresar cmaA1 (Yuan, Lee, et al., 1995) y realizar un *knockout* de cmaA2 (Glickman, Cahill, & Jacobs, 2001). En nuestro caso particular estudiaremos el mecanismo de reacción mediante una dinámica híbrida QM-MM utilizando DFTB como el nivel de teoría cuántico. Al tener el sistema hidratado y realizar dinámica podemos estudiar el rol de la proteína como un todo y el de las moléculas de agua (si lo tuvieran) como así también dilucidar de donde proviene la eficiencia enzimática. Por otra parte también comentaremos las reacciones realizadas en ausencia de cmaA2, es decir, en solución. El sistema proteico fue armado a partir del cristal de cmaA2 (PDBId: 1KPI), este cristal contiene una molécula de S-adenosil-L-homocisteína y un ”inhibidor” análogo del intermediario carbocatiónico (Didecil-Dimetil-amonio) y una molécula de ión carbonato. Utilizando esas coordenadas de partida se construyo un sistema con una olefina en la posición del análogo, SAM en vez de SAH y un ión bicarbonato en vez de carbonato. Dicho sistema fue simulado durante 150ns de dinámica molecular clásica para verificar su estabilidad (ver figura **??**) y obtener fotos de partida adecuadas para poder computar el perfil de energía libre. Para el caso del sistema en agua, el sistema de partida cuanta con una configuración inicial idéntica pero sin la proteína. Para evitar movimientos en el sistema en agua se aplicaron restricciones de distancia para mantener al lípido en su conformación inicial ”cerrada” y mirando al metilo del SAM.

En la dinámica se puede observar que el dominio es bastante rígido y no tiene gran movilidad salvo algunos *loops*. El RMSD se estabiliza a 1.2 Årespecto a la estructura cristalina de partida (Figura **??**) con lo cual se procedió a elegir fotos de partida para la dinámica desde la trayectoria estabilizada.

Formación del carbocatión. Dado que el ataque se puede producir en ambos carbonos de la olefina, se procedió a realizar el ataque a ambos carbonos para dilucidar si existe alguna selectividad intrínseca. En el caso de la olefina utilizada en cmaA2, el carbono C10 es el más cercano al bicarbonato mientras que el C11 se trata del más lejano. Por otra parte se desea

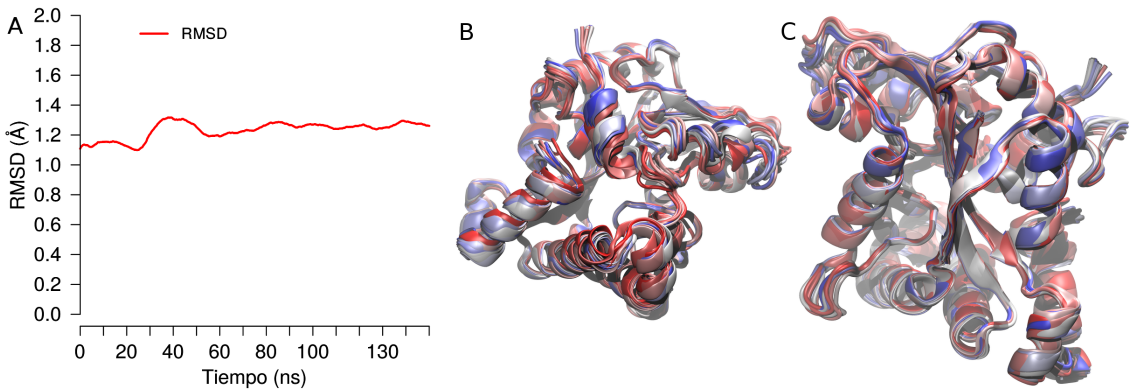


Figura 4.14: **Estabilidad de cmaA2 durante 150ns de dinámica.** A) Calculo de RMSD de los átomos pesados de la cadena principal de cmaA2 respecto a la estructura cristalina de partida (PDBID: 1KPI). B y C) Vista de distintos ángulos de estructuras representativas a lo largo de la trayectoria de 150ns coloreadas desde azul (0ns) a blanco (75ns) a rojo (150ns).

estudiar el rol de los residuos aromáticos en la estabilización del intermediario y el estado de transición, por lo cual también se analizará dicho rol mediante el estudio de la mutante Y30A. Por último se realiza la reacción libre en agua para comprender el rol de la proteína en general y poseer una reacción de referencia contra la cual comparar los resultados de la proteína.

Como se puede observar en la Figura **??**A, la reacción procede como una reacción de ataque nuclefílico (ver figura **??**). El estado de transición se encuentra más cercano a productos que a reactivos, con un valor de la barrera de 14,07 kcal/mol y un ΔG de 2,7 kcal/mol para el caso del ataque desde C10 mientras que la barrera es de 13,95 kcal/mol y un ΔG de -0,31 kcal/mol. En la barrera no existen diferencias significativas entre el ataque a ambos carbonos pero si en el ΔG de la reacción, con una diferencia 3 kcal/mol en favor del ataque al C11. En el perfil de energía libre de la reacción en la mutante Y30A (Figura **??**C) se observa un aumento de la barrera en 2,4 kcal/mol y del ΔG 4,64 kcal/mol respecto al ataque de C10 en la proteína salvaje. Si se analiza que sucede con la reacción en solución, esta ocurre con una barrera de 18,72 kcal/mol y un ΔG de 14,35 kcal/mol. La barrera es mayor y la estabilidad del producto es significativamente menor. (Ver figura **??**)

Las poblaciones de Mulliken son un método para estimar la carga parcial de los átomos. Si se analizan dichas poblaciones para las reacción en presencia de la matriz proteica al realizar el ataque al carbono C10 se puede apreciar que cambian poco hasta acercarse al estado de transición donde se produce un salto y una inversión en la carga entre el SAM (0,7 unidades de carga), y la olefina (0) que luego de la reacción tiene 0,9 unidades de carga, el SAM (ahora

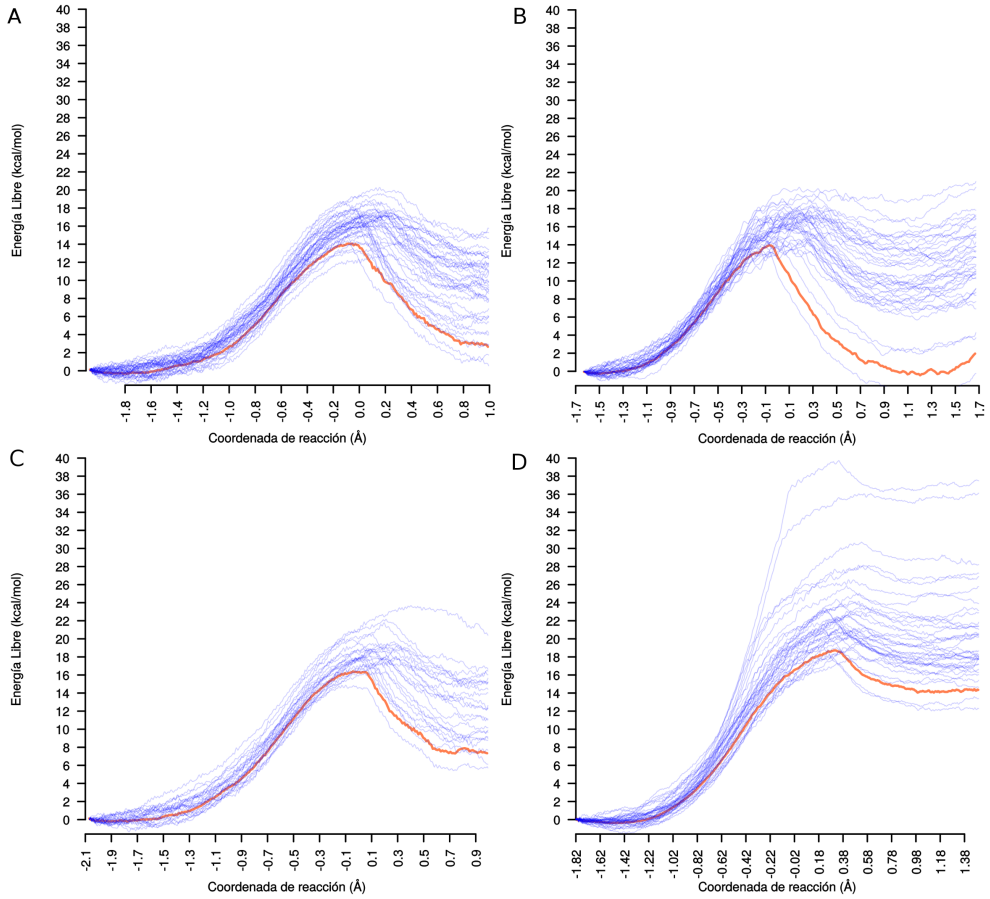


Figura 4.15: **Perfil de energía libre de la reacción de metilación de una olefina en: A) En cmaA2 carbono C10, B) cmaA2 en carbono C11, C) cmaA2 en carbono C10 en mutante Y30A y D) en agua**. En azul se muestran los distintos trabajos y el naranja el perfil de energía libre calculado a partir de la inigualdad de Jarzynski.

SAH) 0 y el metilo conserva 0,1 unidades de carga de las 0,3 que poseía originalmente. (Ver figura **??** A y C). Existe una ligera diferencia de carga en el ataque al carbono C11 dado que el SAM permanece con algo de la carga inicial (0,1 unidades de carga), mientras que el metilo y el lípido se reparten el resto de la carga positiva del sistema (0,1 y 0,8 respectivamente) no pero siguiendo el mismo patrón que en el caso de C10, Figura **??** B. A diferencia de la reacción en la proteína, las cargas presentan un cambio mucho más brusco en el estado de transición alcanzando el lípido valores muy cercanos a 1,0 unidades de carga al final de la reacción. Por otra parte, este cambio no se encuentra tan concertado como en el caso de la reacción en la proteína.

En el caso de las distancias si se observa que el metilo se acerca al carbono a atacar (C10) mientras que la distancia azufre-metilo se mantiene más o menos constante hasta llegar al estado de transición, donde se produce un salto que refleja la transferencia del metilo y la ruptura del enlace azufre-metilo (Ver figura **??**A, B y C). Las distancias siguen una tendencia muy similar

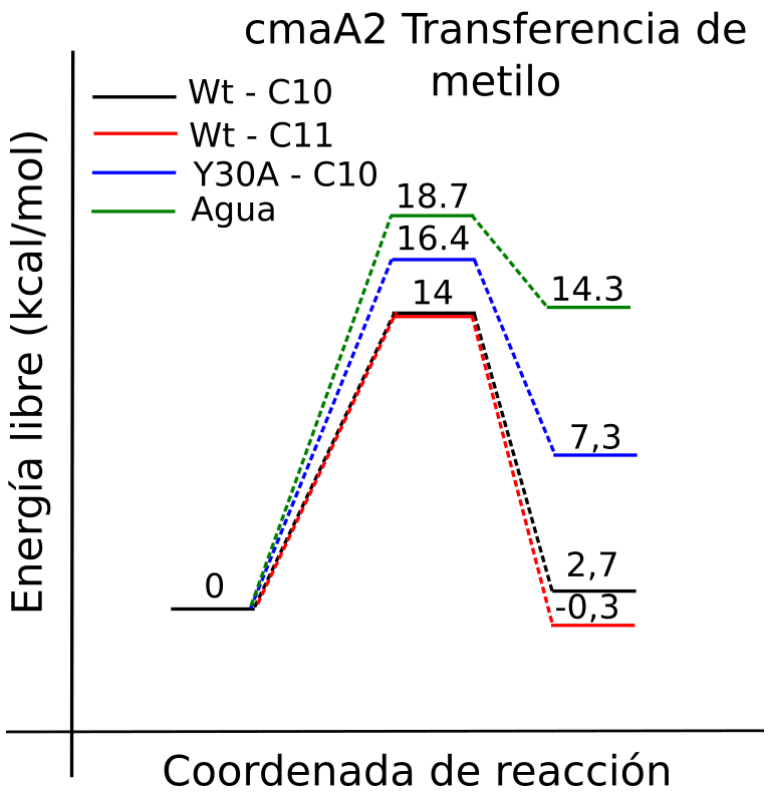


Figura 4.16: **Resumen de los resultados energéticos del primer paso de la reacción de cmaA2.**

a la observada en el proceso realizado en cmaA2 (Ver figura **??**D), con lo cual parece estar adoptando el mismo camino.

La eficiencia catalítica se puede observar claramente en el valor de la barrera (una diferencia de unas 4,77 kcal/mol para el caso de C11 y de 4,65 kcal/mol en el caso de C10) y el ΔG de la reacción (una diferencia de 12 kcal/mol) como así también en la distribución de carga registrada en las poblaciones de Mulliken. La estabilidad en cmaA2, como en otras CMAS, con un abundante cluster de residuos aromáticos (tirosinas y fenilalaninas, en la figura **??** se puede observar la posición de la tirosina 30, por una cuestión de claridad no se muestran el resto de los residuos) que podrían estabilizar el carbocatión mediante interacciones catión-Π. Además, es interesante señalar que el intermediario de reacción general (el producto de este paso) muestra también una marcada estabilización dentro de la proteína.

Resolución del carbocatión. El segundo paso de la reacción en cmaA2 fue realizado en presencia y ausencia de Glu 142, un residuo muy conservado en el subtipo CMA (cmaA1, cmaA2, pcaA, mmaA2) que podría estar involucrado en la reacción de ciclación. Como se mencionó anteriormente en el capítulo, cmaA2 presenta en su sitio activo un ión HCO_3^- que sería el responsable remover un protón al metilo transferido permitiendo la formación del ci-

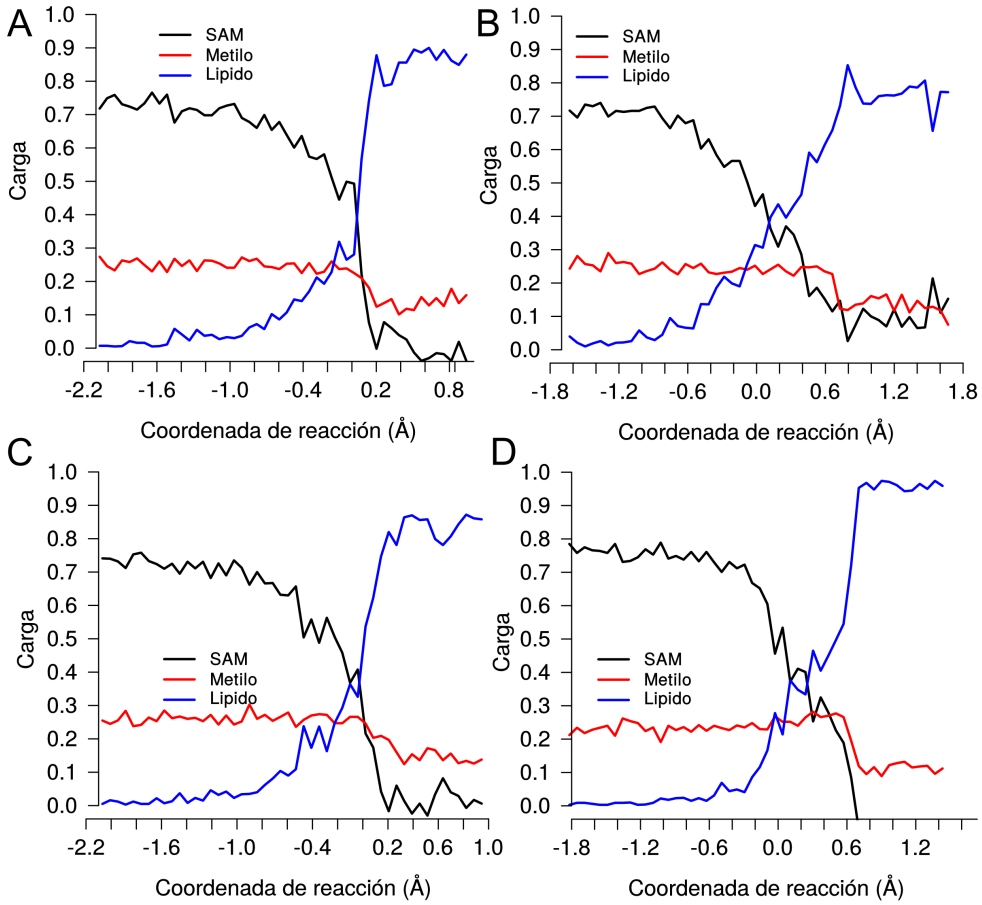


Figura 4.17: **Poblaciones de mulliken de la reacción de metilación de olefinas: A) cmaA2 en C10, B) cmaA2 en C11, C) cmaA2 Y30A en C10 y d) agua.**

clopropano. La presencia de Glu 142 permite que este actúe de forma de base complementaria, al transferirse el protón de HCO3 a Glu 142 y CO3 recibir el protón del metilo que cicla. Un fenómeno que se conoce como *proton relay* por sus siglas en inglés, se puede observar mejor en el esquema presentado en la Figura **??**.

El análisis de los resultados, que se muestran en la figura **??** indican que es posible apreciar que si bien las transferencias en presencia y ausencia de Glu 142 tienen una barrera muy similar y baja (1,35 kcal/mol contra 2,20 kcal/mol) es una diferencia considerable pero dentro del error del método de estimación. Es interesante destacar que el producto final de la reacción en presencia de E142 es bastante más estable (unas 6 kcal/mol) que su contraparte sin E142. Si comparamos el perfil de energía libre en la proteína (en cualquiera de los dos casos) con la reacción en solución, se puede observar que nuevamente la barrera es mucho más alta en comparación a la reacción en cmaA2 (26 kcal/mol contra 2 kcal/mol, ver figura **??**) y el producto final presenta menor estabilidad en agua. La baja barrera en cmaA2 respecto a agua puede explicarse por la presencia de una base fuerte estabilizada en la proteína, mientras que en agua

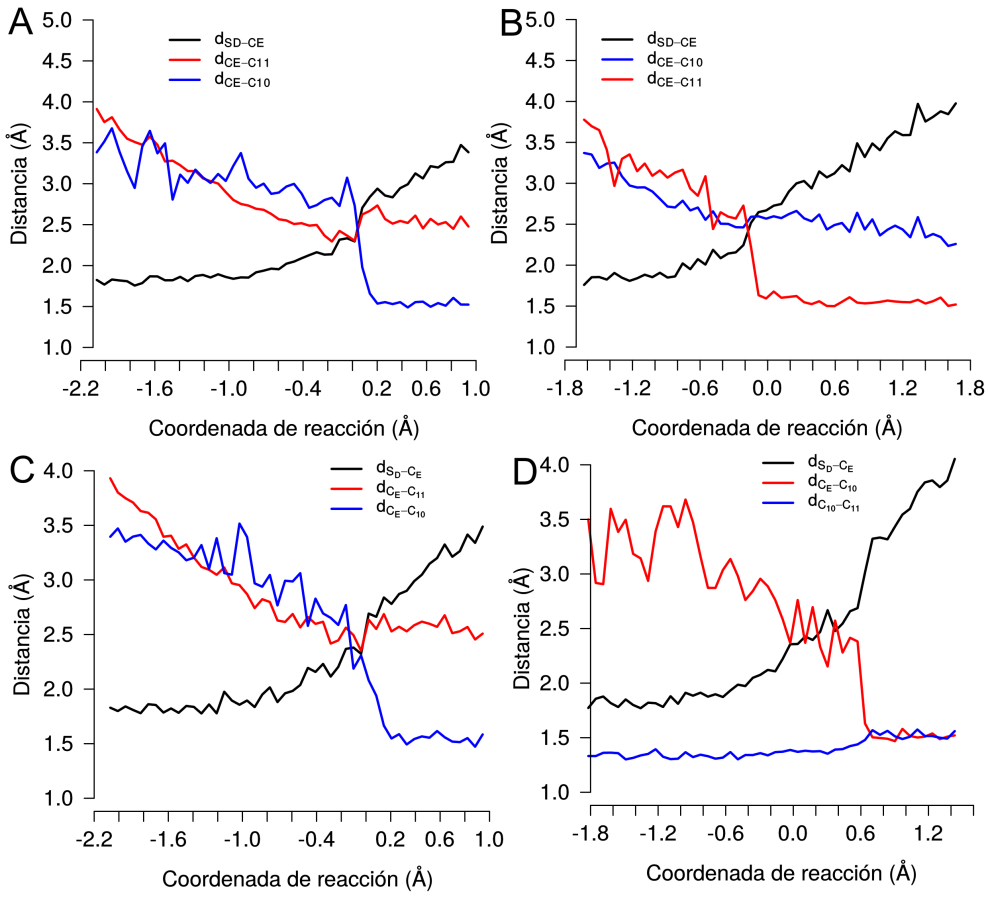


Figura 4.18: **Distancias relevantes en la reacción de metilación de olefinas en:A) cmaA2 en C10, B) cmaA2 en C11, C) cmaA2 Y30A en C10 y d) agua.**

debe ser el solvente mismo él que actúe como base (de forma poco favorable) (Ver figura **??**).

Por el lado de las cargas, en ambos casos la olefina alcanza una carga neutra al final de la reacción. Por el lado de CO3, en presencia de E142 presenta una carga neta negativa cercana a -1 mientras que en ausencia de E142 su valor se acerca mucho más a 0 (0,3-). En el caso de los protones, se puede observar que el protón transferido desde el grupo metilo (H2) casi no tiene carga y que la adquiere al transferirse al CO3. Por otra parte el protón transferido de CO3 a E142 conserva su carga durante todo el proceso. Algo similar sucede en la reacción en ausencia de E142, donde el protón pasa a tener una carga de alrededor de 0,3-0.4. (Ver figura **??** A y B) Para el caso de la reacción en agua se puede encontrar una explicación sobre la diferencia de estabilidad relativa de los productos finales, la formación de un hidronio (H3O+) con la transferencia de carga del, ahora alcano, a la molécula de agua protonada.

Nuevamente las distancias muestran que ambas reacciones siguen el mismo camino, con lo cual estamos evaluando la diferencia en estabilidad provocada por la presencia de E142. En las distancias relevantes de la reacción se puede constatar que la formación del ciclopropano ocurre

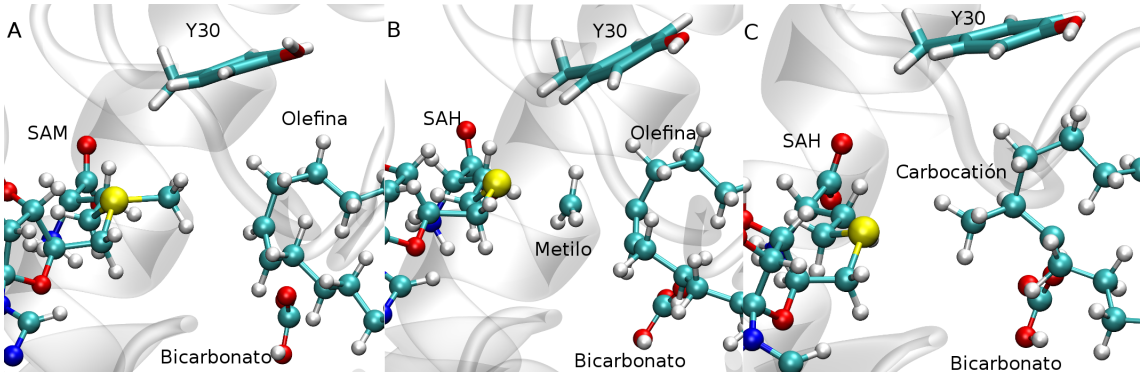


Figura 4.19: **Estructura del A) estado inicial, B) el estado de transición y C) el estado final del primer paso de formación del ciclopropano en cmaA2.**

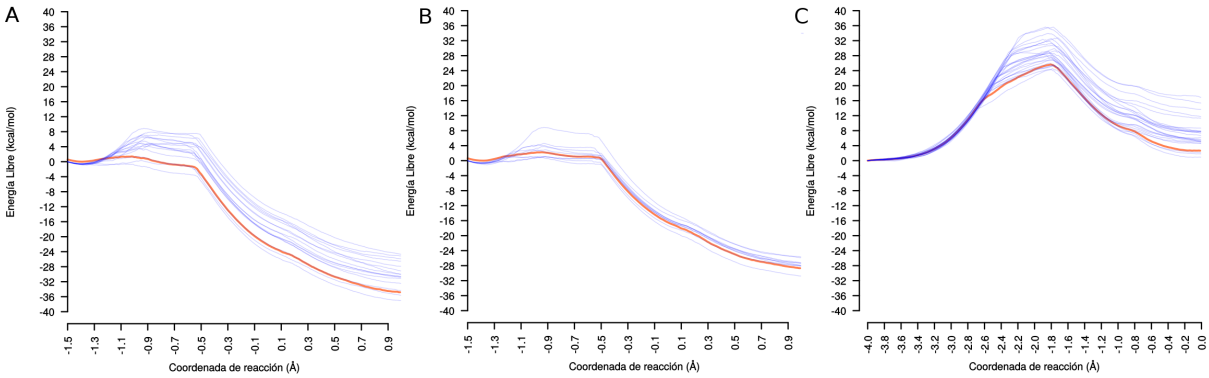


Figura 4.20: **Perfil de energía libre de la reacción de resolución del carbocatión en: A) cmaA2 utilizando E142 como base adicional, B) sin E142 y C) agua.** En azul se muestran los distintos trabajos y el naranja el perfil de energía libre calculado a partir de la inigualdad de Jarzynski.

de forma similar en solución y en cmaA2.

El producto de cmaA2 esta determinado por donde se produce el ataque, si es a C11, el producto es *trans* (el más estable) mientras que si el ataque es por C10 el producto es en *cis* (menos estable y más difícil porque requiere de una mayor reorganización del sitio).

Mecanismo de reacción de mmaA4

Al igual que en cmaA2, para mmaA4 se corrieron dinámicas moleculares clásicas previas para obtener fotos. En este caso se realizó una simulación de 500ns que se muestra bastante estable (ver figura **??**), el único cambio apreciable es la apertura del el loop N-terminal que se puede observar en la Figura **??**A y en el cálculo de de RMSD presentado en **??**.

El primer paso de la reacción en mmaA4 (ver Figura **??** para comprender el mecanismo de reacción y **??** C y D para las distancias relevantes para la coordenada de reacción) ocurre de manera muy similar a cmaA2, con un estado de transición con el metilo plano y una barrera

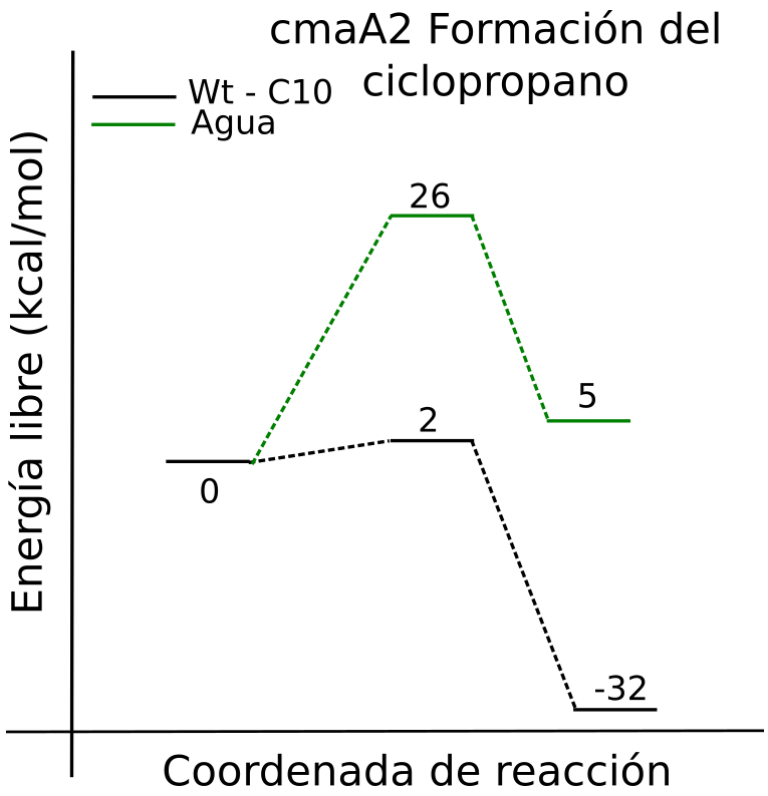


Figura 4.21: Resumen de los resultados energéticos del segundo paso de la reacción de cmaA2.

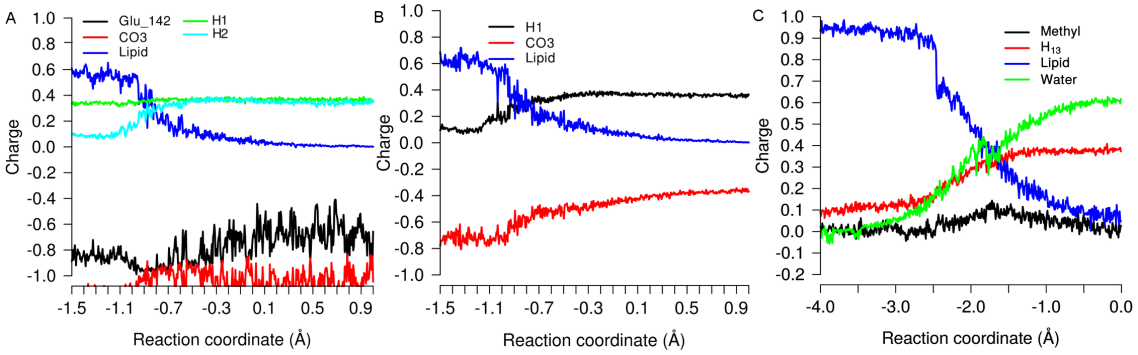


Figura 4.22: **Poblaciones de mulliken de la reacción de resolución del carbocatión en: A) cmaA2 utilizando E142 como base adicional, B) sin E142 y C) agua.**

de valores similares. En este caso se prestó especial atención a la diferencia en barrera en la transferencia a ambos carbonos porque el producto final de esta reacción es quiral (S,S). Como se puede observar en la figura **??** existen una diferencia de 2kcal/mol en la barrera de ambas transferencias, este resultado se encuentra dentro del error del método con lo cual no se puede determinar si se esta favoreciendo o no un estereoisómero sobre el otro.

Si se analizan las cargas y las distancias se observa un comportamiento muy similar al hallado en cmaA2. El estado de transición se encuentra más cercano a productos que a reactivos. La proteína también favorecería la reacción al estabilizar la carga en el estado de transición y el intermediario (Ver figura **??**).

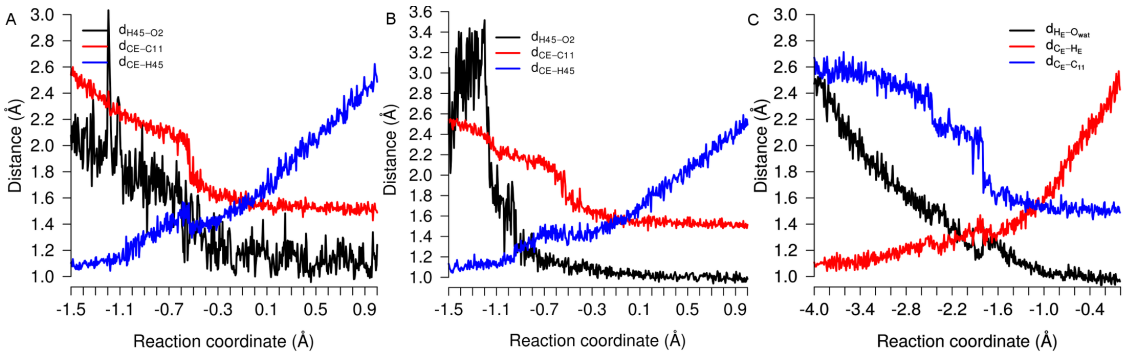


Figura 4.23: **Distancias relevantes de la reacción de resolución del carbocatión en: A) cmaA2 utilizando E142 como base adicional, B) sin E142 y C) agua.**

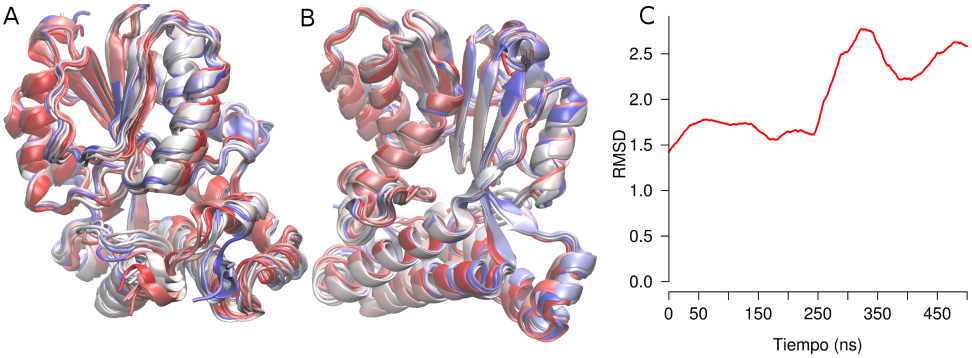


Figura 4.24: **Estabilidad de mmaA4 durante 500ns de dinámica.** A y B) Vista de distintos ángulos de estructuras representativas a lo largo de la trayectoria de 500ns coloreadas desde azul (0ns) a blanco (250ns) a rojo (500ns). C) Calculo de RMSD de los átomos pesados de la cadena principal de mmaA4 respecto a la estructura cristalina de partida (PDBID: 3HA5)

Dado que la reacción ocurre de forma espontanea al producirse una reorganización de los ángulos diedros de la olefina no fue posible realizar ninguna estimación de energía libre, se trató de una simulación de 10ps con un paso de integración de 0.5 fs. en donde en forma espontanea se produce la transferencia del agua y un protón a un glutámico.

A diferencia del segundo paso de cmaA2, con una simple reorganización de la geometría del carbocatión, una rotación del diedro CE-C5-C6-H2 de la molécula permite una transferencia instantánea y sin barrera del agua al carbocatión y su deprotonación por parte de Glu 126. Justamente la disposición de ambos glutámicos como una trampa de moléculas de agua permite que la reacción ocurra sin barrera. La estructura del producto se puede observar en la figura **??**

4.4 Discusión

Los perfiles de energía libre calculados en este capitulo para cmaA2 y mmaA4 permiten determinar en primer lugar que el paso limitante de la reacción es la transferencia de metilo y

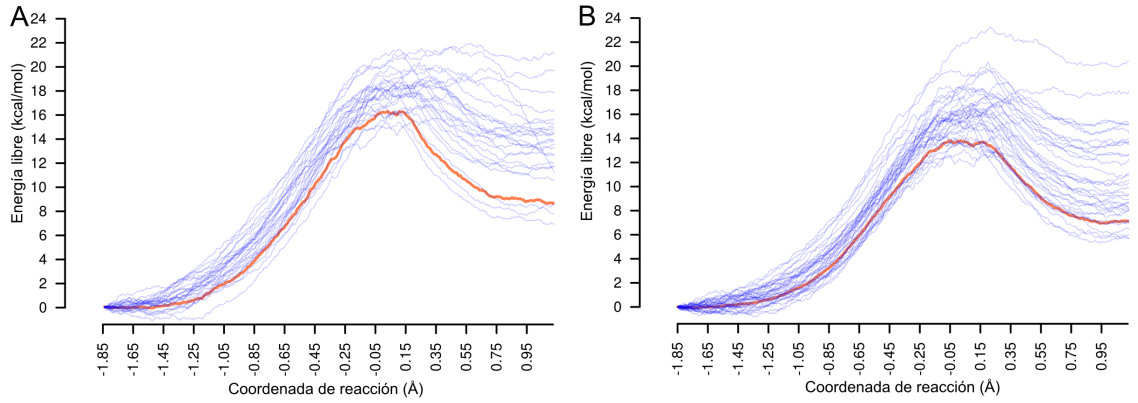


Figura 4.25: **Perfiles de energía libre de la transferencia de metilo y formación del carbocatión al átomo A) C6 y B) C5 de la olefina modelo utilizada.**

formación del intermediario carbocatiónico. Este resultado está de acuerdo con lo determinado para cmaA1 (Liao et al., 2011) mediante optimizaciones restringidas en un sistema reducido aunque con una barrera mucho más alta y también para el homologo de *E. coli* donde es posible, dado el tamaño del sustrato lipídico, realizar determinaciones de constantes cinéticas de actividad. (Courtois, Guérard, et al., 2004; Iwig, Grippe, et al., 2004; Iwig, Uchida, et al., 2005; Courtois, & Ploux, 2005; Guangqi et al., 2013). En los trabajos de Iwig y colaboradores (Iwig, Grippe, et al., 2004; Iwig, Uchida, et al., 2005) se demuestra el rol del bicarbonato mediante un experimento de *scrubbing* o remoción de dioxido de carbono del medio mediante la utilización de PEP-C (Fosfoenolpiruvato carboxilasa) que carboxila el fosfoenolpiruvato, generando oxalacetato que es oxidado por la MDH (Malato deshidrogenasa) a malato permitiendo remover el CO_2) de la solución. Esto permite agregar concentraciones conocidas de sales de bicarbonato ($KHCO_3$, en este caso) y estudiar su rol. Al utilizar su sistema de remoción completo se observa sólo un 3% de la actividad respecto al sistema en presencia de bicarbonato, indicando el rol fundamental que tiene el ión bicarbonato para la actividad ciclopropano sintasa de la CFA de *E. coli*. Por otra parte los autores construyen y ensayan mutantes de la histidina (H266A) y tirosina (Y317F) involucradas en la unión a bicarbonato en las CMAS ocurriendo algo muy similar al experimento de remoción, encontrándose actividades de 2,1% y 0,45% respecto a la proteína salvaje (Iwig, Uchida, et al., 2005). Para probar si el paso limitante de la reacción es la transferencia del metilo y formación del carbocatión (Iwig, Grippe, et al., 2004) realizaron experimentos de efecto calcógeno de la S-adenosil-metionina, reemplazando el átomo de azufre por selenio y telurio. Con dichos experimentos se demuestra que la reacción depende de la identidad del átomo al cual esta unido el metilo al ser transferido, elevándose la transferencia

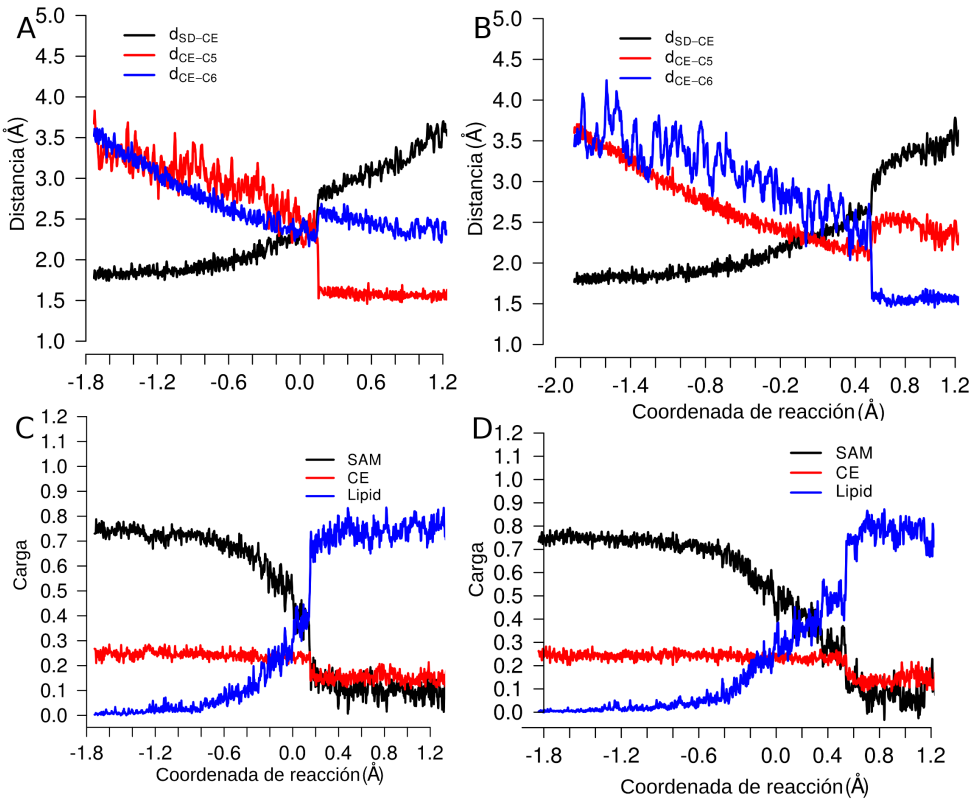


Figura 4.26: **Distancias y cargas de la transferencia de metilo de SAM a una olefina en mmaA4 en sus dos carbonos.** A) Distancia transferencia a C5. B)Cargas transferencia a C5. C) Distancias transferencia a C6. D) Cargas transferencia a C6.

con Selenio y disminuyendo dramáticamente con Telurio. (Iwig, Grippe, et al., 2004)

Al comparar la reacción realizada por cmaA2 y mmaA4 respecto a la computada en agua se observa una diferencia de unas 4 kcal/mol del estado de transición y 8 kcal/mol respecto al producto, siendo mucho más estables dentro de las CMAS que fuera. Todas las CMAS presentan en el sitio de unión un grupo de aminoacidos aromáticos (Tirosinas y fenilalaninas) en gran proximidad con el SAM/SAH y la olefina. La forma más fácil de estabilizar a un carbocatión sería tener residuos ácidos en la proximidad del carbocatión pero esto también conllevaría una potencial reactividad del anión en cuestión con el carbocatión a estabilizar. Esta misma estabilidad se puede lograr utilizando el momento cuadrupolar de un residuo aromático como son las tirosinas, fenilalaninas o triptofanos. Si bien se encuentran residuos ácidos o aniones en las cercanías del carbocatión (los glutamicos de mmaA3/4 y el bicarbonato en el resto de las CMAS) no se encuentran tan cerca como el *cluster* de aromáticos presente en todas las CMAS. Esta conservación puede estar hablando de un rol funcional a la hora de la catálisis (en este caso electroestática) llevada a cabo por estas enzimas. (Thibodeaux, Chang, & Liu, 2011; Huang et al., 2002) El rol de los aromáticos queda demostrado en los cálculos del perfil

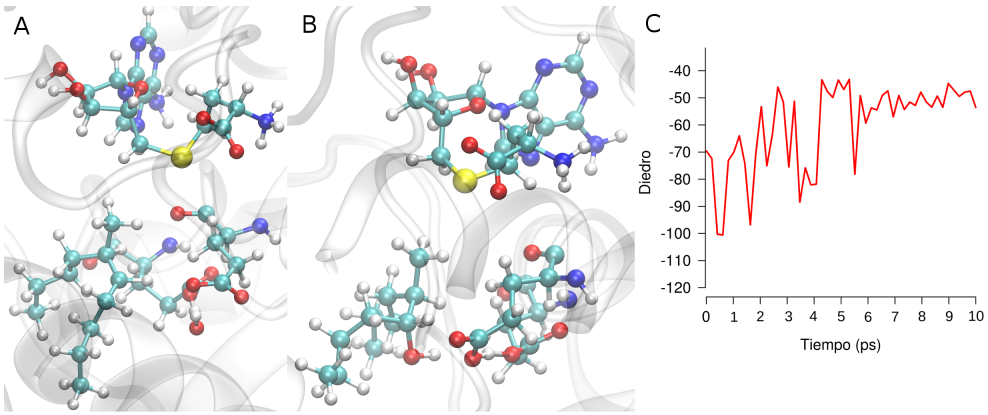


Figura 4.27: **Estructura del producto de reacción llevada a cabo por mmaA4.**

de energía libre del primer paso para la mutante Y30A (el aminoácido aromático más cercano a la posición de la olefina) de cmaA2 ya que al realizar dicho perfil se observa una barrea 2,4 kcal/mol superior a la proteína salvaje mientras que el ΔG aumenta en 4,64 kcal/mol.

El segundo paso ocurre catalizado en ambas enzimas. En el caso de cmaA2 la reacción de ciclación ocurre con una barrera baja (alrededor de 2kcal/mol) mientras que en solución tiene una barrera de 26 kcal/mol. Ciertamente la proteína esta actuando como base a través de la unión del bicarbonato y el glutámico (que favorece la reacción). En mmaA4 la reacción directamente ocurre sin barrera al producirse un reorientación del lípido, ciertamente ayudada en gran medida por el posicionamiento correcto de la molécula de agua que ataca al carbocation por parte de los dos glutámicos de mmaA4, que actúan como base al sustraer el protón del agua.

Con los resultados del modelado, alineamientos estructurales y de secuencia y los perfiles de energía libre de cmaA2 y mmaA4 es posible establecer un modelo que explique la selectividad en las diferentes enzimas que conforman la familia. La diferencia entre mmaA3 y mmaA4 respecto de la selectividad podrían explicarse en términos de la entrada diferencial de aguas al sitio activo, producto del cierre y estabilización posterior a la entrada de sustratos en mmaA3 (ver figura **??**A). Al no existir un cierre el agua en mmaA4 puede entrar y salir con menor restricción de la proteína, siendo este paso fundamental para que se produzca la reacción y se realice un recambio del protón que ha adquirido la proteína al realizar la sustracción. En mmaA3 la falta de aguas en el sitio activo podría provocar un cambio en el pKa de los glutámicos que favorezca enormemente el ataque por parte de estos residuos al grupo alcohol agregado por mmaA4.

Para el caso de la selectividad en las CMAS más tradicionales (cmaA1-2, mmaA2 y pcaA),

la presencia de bicarbonato y glutámico parece ser crítica para su actividad. Sin el bicarbonato no tendrían la basicidad suficiente para secuestrar un protón del metilo y cerrar el ciclopropano, siendo este el paso limitante del segundo paso de la reacción. MmaA1 realiza la reacción de metilación y posterior formación de un doble enlace en *trans*. La principal diferencia de mmaA1 frente al resto de las CMAS es la sustitución de un glutámico por un aspártico, que tiene una cadena más corta. Es posible que este hecho impida la retención de aguas observada en mmaA4 o que se favorezca la forma protonada del ácido como podría ocurrir en mmaA3.

Respecto a umaA, ufaA1 y rv3720. Desde un punto de vista de secuencia, umaA tiene una mayor similitud a mmaA1 que al resto de las CMAS, elemento que se puede observar en la Figura **??**. Sin embargo presenta una configuración del sitio activo idéntica a cmaA2 con lo cual cabría esperar que umaA tenga una actividad similar a esta u otra de las *Cyclpropane synthases*. Experimentalmente existe un estudio del grupo de Daffé (Laval et al., 2008) en donde las mutantes de umaA en *Mtb H37Rv* no presenta cambios en su perfil de ácidos micólicos, cosa que si sucede con el homólogo en *M. smegmatis*. Algo similar ocurre con cmaA1, que si bien tiene la configuración del sitio activo de cmaA2, experimentos de *knockout* no muestran cambios en el perfil de producción de ácidos micólicos, con lo cual su rol *in vivo* resulta desconocido. UfaA1, presuntamente, estaría involucrada en la síntesis de ácido tubercuesteárico (Ácido 10-Metiloctadecanoico) (Meena, & Kolattukudy, 2013). Dado que la reacción desde el carbocatión requiere de poder reductor, los autores proponen que es brindado por NADPH. Si bien es posible, dado que ufaA1 posee un dominio N-terminal con 140 aminoácidos más que el resto de las CMAS con lo cual con ese dominio extra podría estar uniendo NADPH pero se requieren de más estudios para determinar fehacientemente la actividad de ufaA1.

Estos resultados podrían ser validados realizando mutantes en la enzima de *E. coli* en donde resulta relativamente fácil realizar experimentos de actividad enzimática. Principalmente se podría evaluar el rol del agrupamiento de aromáticos en la estabilización del estado de transición 1 y el intermediario, dado que estudios similares se han realizado con la mutante puntual G236E (posición equivalente al primer glutámico del sitio activo de mmaA4), no produciéndose el producto deseado (metil-alcohol) sino un producto similar al observado en mmaA1.(Guangqi et al., 2013). Justamente como se señaló arriba, probablemente la entrada diferencial de aguas entre mmaA4, que se ve claramente favorecida por la falta de estructura del N-terminal, sea el

	Sitio activo	**Helice η1**	**Helice ηX**	**Estructura**	**Producto**
Proteina	**Bicarbonato**	**Glutámico**	**Histidina**	**helice ηX**	**Propuesto**
cmaA1	Si	Glu	His	Hélice	Cis-ciclopropano
cmaA2	Si	Glu	His	Hélice	Cis/Trans-ciclopropano
pcaA	Si	Glu	His	Hélice	Cis-ciclopropano
mmaA1	Glu	Asp	His	Hélice	Trans-Metil-olefina
mmaA2	Si	Glu	His	Hélice	Cis-ciclopropano
mmaA3	Glu	Glu	His	Hélice	Metil-eter
mmaA4	Glu	Glu	His	Desestruct.	Metil-alcohol
umaA	Si	Glu	Ala	Hélice	¿Cis-ciclopropano?
ufaA1	Glu	Glu	His	Hélice	Ac. 10-metiloctadecaoico
rv3720	Si	Glu	His	Hélice	Desconocida

Tabla 4.1: Resumen de las diferencias entre las distintas proteínas de la familia CMAS de *Mtb*

paso determinante en la actividad de mmaA4. Al no existir una estructura y poseer una baja identidad de secuencia (por debajo del 34%) no permite sacar conclusiones de tipo estructural pero si se alinean las dos secuencias se puede concluir que las diferencias se encuentran, como en el caso de ufaA1 y rv3720, en el dominio N-terminal desconocido.

Si bien desde un punto de vista estructural y de secuencia hemos podido resolver ciertas cuestiones de la actividad específica de algunas de las CMAS, existen preguntas abiertas como comprender la selectividad entre el sitio distal y proximal del ácido micólico, probablemente relacionado tanto al largo del sustrato como a la interacción proteína-proteína que media la interacción (dado que el micólico no se encuentra libre en solución, sino unido covalentemente a una proteína *carrier* -ACP-). Otra pregunta abierta y con relación a la selectividad entre sitios es si las reacciones de modificación ocurren durante la síntesis o luego de ella, ya que existen experimentos de doble híbrido que muestran interacciones entre proteínas del complejo FAS II y las CMAS. (Cantaloube et al., 2011). Por último es interesante destacar, cuestión que ampliaremos en el siguiente capítulo, que el sitio activo de las CMAS presenta características interesantes para el desarrollo de inhibidores con propiedades polifarmacológicas, dado que el mayor efecto in vivo se observa al eliminar a varios de los componentes del sistema (por ejemplo, mmaA2 y cmaA2 presentan cierta redundancia y solapamiento).

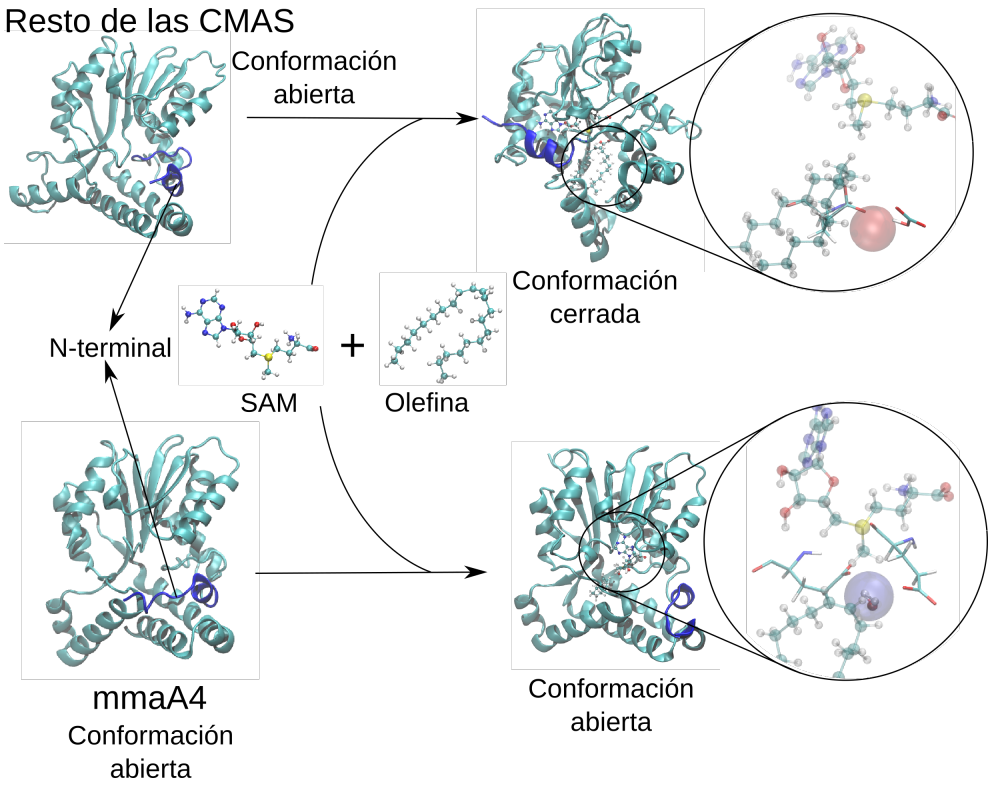


Figura 4.28: **Modelo propuesto del funcionamiento de las CMAS.** Arriba, las CMAS en general, abajo mmaA4. Al producirse la unión de los sustratos en mmaA4 no se produce un cierre del N-terminal mientras que en el resto de los miembros de la familia. La ampliación muestra que mmaA4 posee un sitio de aguas entre los dos glutámicos (un lugar donde es más probable encontrar aguas que en el seno del solvente) mientras que en cmaA2 este es mucho más débil.

Bibliografía

Barkan, Daniel et al. (2010). "Redundant function of cmaA2 and mmaA2 in Mycobacterium tuberculosis cis cyclopropanation of oxygenated mycolates". In: *Journal of bacteriology* 192.14, pp. 3661–3668.

Behr, Marcel A et al. (2000). "A point mutation in the mma3 gene is responsible for impaired methoxymycolic acid production in Mycobacterium bovis BCG strains obtained after 1927". In: *Journal of bacteriology* 182.12, pp. 3394–3399.

Cantaloube, Sylvain et al. (2011). "The Mycobacterium tuberculosis FAS-II dehydratases and methyltransferases define the specificity of the mycolic acid elongation complexes". In: *PloS one* 6.12, e29564.

Case, DA et al. (2014). "Amber 14". In:

Courtois, Fabienne, Christine Guérard, et al. (2004). "Escherichia coli cyclopropane fatty acid synthase". In: *European Journal of Biochemistry* 271.23-24, pp. 4769–4778.

Courtois, Fabienne, & Olivier Ploux (2005). "Escherichia coli cyclopropane fatty acid synthase: is a bound bicarbonate ion the active-site base?" In: *Biochemistry* 44.41, pp. 13583–13590.

Defelipe, LA et al. (2011). "Juvenile hormone synthesis:"esterify then epoxidize" or "epoxidize then esterify"? Insights from the structural characterization of juvenile hormone acid methyltransferase". In: *Insect biochemistry and molecular biology* 41.4, pp. 228–235.

Glickman, Michael S (2003). "The mmaA2 gene of Mycobacterium tuberculosis encodes the distal cyclopropane synthase of the α-mycolic acid". In: *Journal of Biological Chemistry* 278.10, pp. 7844–7849.

Glickman, Michael S, Sean M Cahill, & William R Jacobs (2001). "The Mycobacterium tuberculosis cmaA2 gene encodes a mycolic acid trans-cyclopropane synthetase". In: *Journal of Biological Chemistry* 276.3, pp. 2228–2233.

Guangqi, E et al. (2013). “An active site mutant of Escherichia coli cyclopropane fatty acid synthase forms new non-natural fatty acids providing insights on the mechanism of the enzymatic reaction”. In: *Biochimie* 95.12, pp. 2336–2344.

Huang, Chih-chin et al. (2002). “Crystal structures of mycolic acid cyclopropane synthases fromMycobacterium tuberculosis”. In: *Journal of Biological Chemistry* 277.13, pp. 11559–11569.

Iwig, David F, Anthony T Grippe, et al. (2004). “Isotope and elemental effects indicate a rate-limiting methyl transfer as the initial step in the reaction catalyzed by Escherichia coli cyclopropane fatty acid synthase”. In: *Biochemistry* 43.42, pp. 13510–13524.

Iwig, David F, Akira Uchida, et al. (2005). “The activity of Escherichia coli cyclopropane fatty acid synthase depends on the presence of bicarbonate”. In: *Journal of the American Chemical Society* 127.33, pp. 11612–11613.

Laval, Françoise et al. (2008). “Investigating the Function of the Putative Mycolic Acid Methyltransferase UmaA DIVERGENCE BETWEEN THE MYCOBACTERIUM SMEGMATIS AND MYCOBACTERIUM TUBERCULOSIS PROTEINS”. In: *Journal of Biological Chemistry* 283.3, pp. 1419–1427.

Liao, Rong-Zhen et al. (2011). “Mechanism of mycolic acid cyclopropane synthase: a theoretical study”. In: *Biochemistry* 50.9, pp. 1505–1513.

Marrakchi, Hedia, Marie-Antoinette Lanéelle, & Mamadou Daffé (2014). “Mycolic acids: structures, biosynthesis, and beyond”. In: *Chemistry & biology* 21.1, pp. 67–85.

Martin, Jennifer L, & Fiona M McMillan (2002). “SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold”. In: *Current opinion in structural biology* 12.6, pp. 783–793.

Meena, Laxman S, & Pappachan E Kolattukudy (2013). “Expression and characterization of Rv0447c product, potentially the methyltransferase involved in tuberculostearic acid biosynthesis in Mycobacterium tuberculosis”. In: *Biotechnology and applied biochemistry* 60.4, pp. 412–416.

Poirot, Olivier, Eamonn O’Toole, & Cedric Notredame (2003). “Tcoffee@ igs: a web server for computing, evaluating and combining multiple sequence alignments”. In: *Nucleic acids research* 31.13, pp. 3503–3506.

Radusky, Leandro et al. (2014). "TuberQ: a Mycobacterium tuberculosis protein druggability database". In: *Database* 2014, bau035.

Ramirez, Claudia L et al. (2014). "Improving Efficiency in SMD Simulations Through a Hybrid Differential Relaxation Algorithm". In: *Journal of Chemical Theory and Computation* 10.10, pp. 4609–4617.

Thibodeaux, Christopher J, Wei-chen Chang, & Hung-wen Liu (2011). "Enzymatic chemistry of cyclopropane, epoxide, and aziridine biosynthesis". In: *Chemical reviews* 112.3, pp. 1681–1709.

Wang, Junmei et al. (2004). "Development and testing of a general amber force field". In: *Journal of computational chemistry* 25.9, pp. 1157–1174.

Yuan, Ying, & Clifton E Barry (1996). "A common mechanism for the biosynthesis of methoxy and cyclopropyl mycolic acids in Mycobacterium tuberculosis". In: *Proceedings of the National Academy of Sciences* 93.23, pp. 12828–12833.

Yuan, Ying, Deborah C Crane, et al. (1997). "MMAS-1, the branch point between cis-and trans-cyclopropane-containing oxygenated mycolates in Mycobacterium tuberculosis". In: *Journal of Biological Chemistry* 272.15, pp. 10041–10049.

Yuan, Ying, Richard E Lee, et al. (1995). "Identification of a gene involved in the biosynthesis of cyclopropanated mycolic acids in Mycobacterium tuberculosis". In: *Proceedings of the National Academy of Sciences* 92.14, pp. 6630–6634.

Capítulo 5

Búsqueda virtual de inhibidores de CMAS

5.1 Introducción

La búsqueda y desarrollo de fármacos es un problema de intensa investigación que se encuentra en la frontera del conocimiento. Existen distintas estrategias para abordar el problema, que involucran técnicas *in vitro, in sillico, in vivo* o una combinación de alguna de ellas, cuya explicación se ha abordado en la introducción de esta tesis. En este capítulo utilizaremos una de las CMAS descriptas en el capitulo anterior, umaA, como blanco terapéutico. Su elección se basa en que la vía de síntesis de ácidos micólicos se encuentra actualmente validada como blanco terapéutico en modelos infecciosos (Barkan et al., 2012) y han habido desarrollos promisorios con fármacos de baja potencia para atacar a las CMAS. (Alahari et al., 2007) Aún así, el modo de acción molecular de los mismos es desconocido impidiendo un mejoramiento racional de estos compuestos.

En particular, umaA resulta de interés por tres razones: en primer lugar, al ser las CMAS enzimas muy similares desde el punto de vista estructural en su sitio activo, cualquier enzima que pertenezca a esta familia y que posea las características básicas de su sitio activo puede servir como base para estudios de inhibición y unión en la familia. En segundo lugar, pertenece a la subfamilia de CMAS que unen bicarbonato en su sitio activo, que puede actuar como potencial sitio farmacofórico, factor clave y que detallaremos más adelante en este capítulo. Por último, desde una concepción básica es una de las proteínas menos estudiadas de la familia, razón por la cual resulta novedosa *per se*.

El objetivo de este capitulo, en este contexto comprende utilizar la información acerca del

funcionamiento de las CMAS para proponer moléculas que sean capaces de inhibir la actividad metiltransfersa de estas enzimas. Para ello utilizaremos *docking* para realizar una primera priorización de compuestos y dinámica molecular clásica, para comprender en un contexto más realista el modo de unión de los las moléculas halladas.

5.2 Materiales y métodos

Armado de la base de datos de compuestos

Para el armado de la base de datos de compuestos se realizaron búsquedas en la base de datos pública ZINC (ZINC Is Not Commercial) (Irwin et al., 2012). Se utilizó ZINC por dos razones; la primera es que ZINC es una base de datos pública; la segunda es que se encuentran depositados en ZINC conjuntos curados de compuestos que son accesibles a través de una serie de proveedores. Actualmente ZINC contiene 35 millones de compuestos disponibles para realizar búsquedas virtuales. Para la búsqueda realizada en ZINC se utilizaron 5 fragmentos miméticos de la unión de bicarbonato HCO_3 (ácido tartárico, carbamato, tipo glicina, 5-hidroxi-2(5H)furanona y urea, ver figura **??**) obteniendo unos 30.000 compuestos aproximadamente. Como el estado de protonación y la relevancia de cada estado en cada una de las 30.000 moléculas es desconocida se utilizó un programa capaz de calcular de forma empírica estos datos, a través de aproximaciones en fragmentos conocidos y como grupos cercanos afectan el pKa. LigPrep (Schrödinger, 2011) fue utilizado para generar las especies químicas relevantes (tautómeros, estereoisómeros y micro-especies) a pH 7 como así también los confórmeros tridimensionales de mínima energía.

Docking en umaA

El receptor (umaA, un modelo generado a partir del *pipeline* descrito en la sección de métodos bioinformáticos) fue protonado con tLeap (AMBER) (D. Case et al., 2014) y sometido a una ronda de minimización con el programa SANDER (D. Case et al., 2014) (2000 pasos en vacío, con un corte en las interacciones electrostaticas de 10 Å) para acomodar las posiciones de las cadenas laterales. Se convirtió el archivo de salida al formato MOL2 con OpenBabel 2.3.1 (OLBoyle et al., 2011) que realiza la asignación de tipos de átomos a Tripos 5.2 y el cálculo

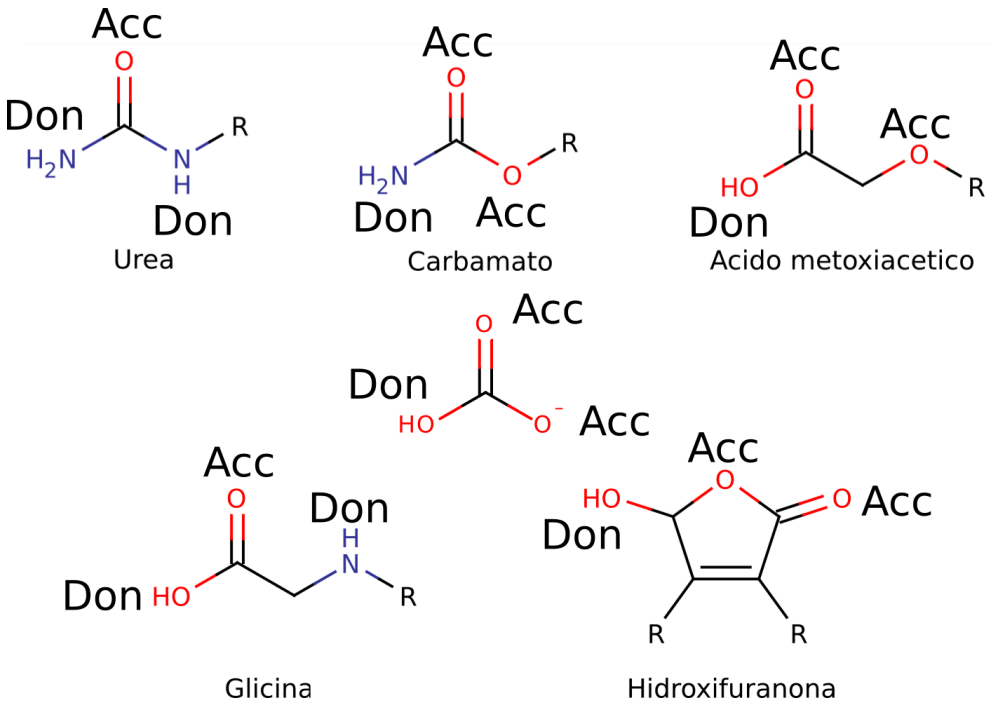


Figura 5.1: **Fragmentos mimeticos del modo de unión de bicarbonato**. En la figura se encuentran señalados el perfil de interacciones de puente hidrógeno de cada fragmento y del anión bicarbonato. Don: Donor de puente hidrogeno, Acc:Aceptor de puente hidrógeno

de las cargas de Garsteiger. El sitio de unión se definió mediante el método de ligando de referencia (Li, Chen, & Weng, 2003; Ruiz-Carmona et al., 2014) utilizando las posiciones de los ligandos cristalizados de cmaA1, previamente alineada estructuralmente con el modelo de umaA. El protocolo de *docking* en si fue explicado en secciones anteriores. Se realizaron 50 determinaciones independientes para cada compuesto eligiendo la pose de menor puntaje total de cada compuesto (energía). La comparación entre compuestos se realiza utilizando el puntaje de interacción S_{inter}, por lo cual los compuestos resultantes fueron ordenados de menor a mayor S_{inter} y luego se analizaron visualmente los primeros 1000 complejos obtenidos.

5.2.1 Dinámica molecular clásica de umaA con ligandos

La dinámica molecular de complejos proteína-ligando permite comprender con mejor detalle la interacción y al mismo tiempo obtener una estimación de la energía libre de interacción. Los parámetros atómicos de los compuestos elegidos para el campo de fuerza AMBER fueron computados. Brevemente, a partir de las poses surgidas del *docking* se realizó una minimización y posterior cálculo de las cargas RESP utilizando AM1-bcc. La asignación de tipos de átomos fue realizada con antechamber con el campo de fuerza GAFF (*General Amber Force Field*) que

resulta en una buena aproximación para los parámetros de enlace. Los parámetros se encuentran disponibles a pedido.

Los sistemas fueron protonados y solvatados en una caja de aguas TIP3P de 10 Åde radio desde el borde de la proteína. Hasta llevar los sistemas a régimen, se colocaron restricciones armónicas en los carbonos alfa (100kcal/mol para la minimización, 10kcal/mol para la termalización y 1 kcal/mol para la equilibración). Adicionalmente se utilizaron restricciones para conservar las distancias de las interacciones de puente hidrógeno encontradas en los experimentos de *docking*. Las distancias entre donor y aceptor de puente hidrógeno se mantuvieron a, como mucho, 3.5 Åcon una penalidad de 50 kcal/mol*$Å^2$.

Una vez armados se procedió a minimizar los sistemas con SANDER (D. Case et al., 2014) durante 10000 pasos, en condiciones periódicas de borde a volumen constante y con un corte de las interacciones de largo alcance a 10 Å. Luego se procedió a un calentado suave del sistema desde 10K a 300K durante 100ps con el termostato de Berendsen (constante de tiempo 2 ps). Por último se procedió a equilibrar el sistema a 300K y 1 bar (Barostato de Berendsen, constante de tiempo 2ps) realizando simulaciones en el ensamble NPT durante 100ps. Se corrieron en tandem 20ns de dinámica molecular en NPT a 300K con las restricciones, para permitir relajar al complejo, y otros 20ns sin ningún tipo de restricciones. En todos los caso se utilizó un paso de integración de 2 fs y el algoritmo SHAKE (Miyamoto, & Kollman, 1992) para mantener los enlaces X-H en su distancia de equilibrio.

MM-PBSA de umaA con ligandos

Las estimaciones de energía libre de unión de los compuestos se realizó mediante MM-(GB)PBSA. Se tomaron 1000 fotos de las trayectorias distanciadas en 10ps entre cada una (para eliminar efectos de correlación) y se realizó el cómputo con una concentración de sales de 0.1M . Para GB se utilizó el modelo desarrollado por Onufriev y colaboradores (Hawkins, Cramer, & Truhlar, 1996; Onufriev, Bashford, & D. A. Case, 2004) mientras que para PB se utilizó el desarrollado por Luo y colaboradores (Luo, David, & Gilson, 2002).

5.3 Resultados

5.3.1 Elección de umaA y el farmacóforo

Como se explicó en la introducción la elección de umaA se encuentra fundamentada, principalmente, en la importancia de la vía de síntesis de ácido micólico y de la importancia de las modificaciones que las CMAS realizan sobre los mismos para el mecanismo infectivo de *Mtb*. Por otra parte, tanto umaA como pcaA, cmaA1-2 y mmaA2 poseen un sitio de unión a bicarbonato que puede ser utilizado como potencial sitio farmacofórico, es decir buscar que los potenciales fármacos tengan un grupo que sea mimético del bicarbonato. La utilización de farmacóforos brinda mayor certeza a la hora de búsqueda de modos de unión de drogas, ya que, al conocer el modo de unión de un fragmento ancla puede ayudar a la búsqueda del modo de unión de toda una molécula. Por esta razón se lo eligió, dado que es una característica importante y, por demás, necesaria para la actividad de las CMAS del subgrupo a estudiar. Como se trata de algo común en este grupo, es de esperar que los compuestos que se encuentren tendrán actividades similares para con el resto de la familia. El sitio de unión a bicarbonato se puede observar en la figura **??**. Este sitio se encuentra formado por una histidina (que puede actuar como donor puente hidrógeno, es decir se encuentra protonada en el nitrógeno epsilon), el grupo amida de una cadena principal de una cisteína (como donor de puente hidrógeno), una tirosina (actúa como donor de puente hidrógeno), alternativamente puede formar interacciones de puente hidrógeno un glutámico que actúa como aceptor. Es decir, el bicarbonato cuenta con dos átomos de oxigeno que actúan como aceptores de puente hidrógeno y otro que puede actuar como donor de puente hidrógeno.

5.3.2 Armado y procesado de la base de datos

Dado que se quiere buscar compuestos que desplacen al bicarbonato, se necesitan fragmentos que puedan imitar el modo de unión descrito para el mismo en la sección anterior. Para ello se utilizaron 5 fragmentos distintos (descritos en la Figura **??**) basados en urea, en carbamato, en acido metoxiacético, en glicina y en hidroxifuranona que pueden imitar el modo de unión debido a su capacidad de realizar y aceptar puentes hidrógeno. En si, el ácido metoxiacético, el carbamato y la hidroxifuranona tienen exactamente la misma configuración de aceptores (2)

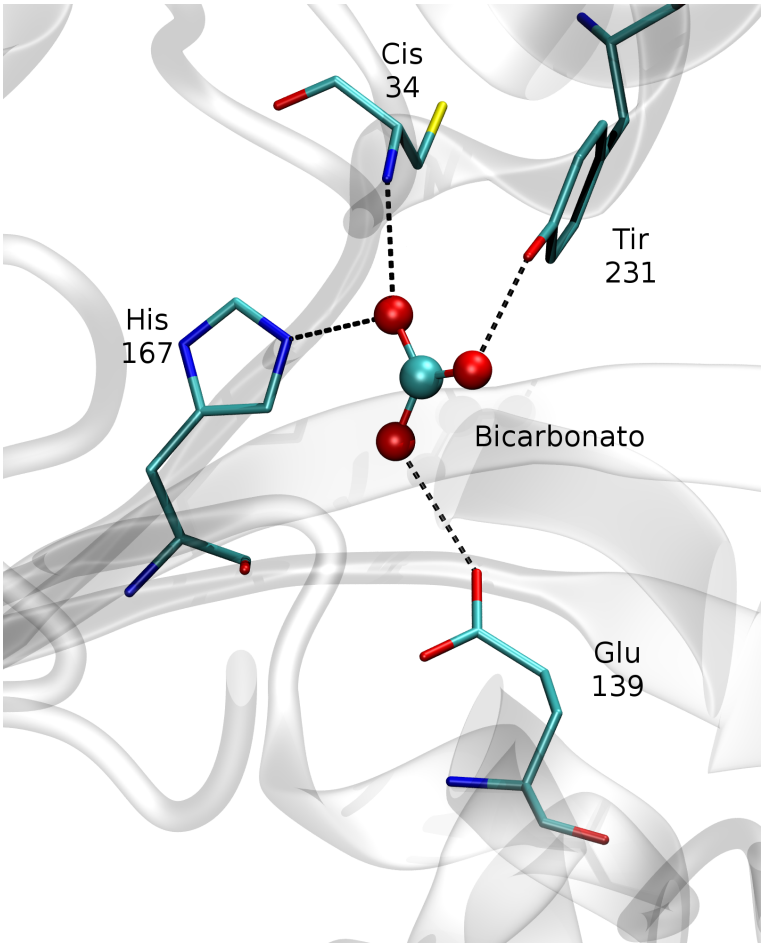


Figura 5.2: **Sitio de unión de bicarbonato de umaA.** La numeración corresponde a umaA mientras que el bicarbonato proviene de una superposición de una estructura de cmaA2 (PDBId: 1KPH) con umaA. Las lineas punteadas corresponden a interacciones de puente hidrógeno. El anión bicarbonato se encuentra unido mediante cuatro puentes hidrógeno actuando en tres casos como aceptor (His(NE) 167, Cis 34 y Tir 231) y en un caso como donor (Glu 139).

y donores (1) que el anión bicarbonato. Por otra parte, la urea y la glicina poseen una configuración distinta privilegiando los donores (2) a los aceptores (1) Estos fragmentos son la base para realizar búsquedas en la base de datos de compuestos ZINC y como el objetivo es, a la larga, poder probarlos *in vitro* es necesario que sean comprables y de fácil disponibilidad. Dichas búsquedas arrojaron 18883 compuestos en base a urea, 141 compuestos en base a carbamato, 4782 compuestos en base a ácido metoxiacético, 1926 compuestos en base a glicina y 164 basados en hidroxifuranona, un total de 25896 compuestos. A partir de allí, utilizando el motor de preparación de compuestos LigPrep, se computaron los correspondientes micro-especies (relevantes al 1% a pH 7), tautómeros, estereoisómeros y confórmeros obteniendo unas 59435 estructuras distintas. Esta es la biblioteca de compuestos que se sometió a un experimento de búsqueda virtual con rDock.

5.3.3 *Docking*

El *docking* se ejecutó con rDock realizado 50 corridas independientes para cada estructura, con un *restraint* farmacofórico para que las poses obtenidas se encuentren enriquecidas en interactores de características similares al bicarbonato. Para ello se aplicaron 2 *restraints* farmacoforicos, uno de aceptor de puente hidrógeno(para interaccionar con la histidina 167 y la cadena principal de cisteína 34) y otro de donor de puente hidrógeno (para interaccionar con el glutámico 139). Solo se conservaron los compuestos que cumplieran con ese *restraint* farmacofórico y tengan un puntaje de interacción menor a -25 unidades (el puntaje es un símil delta G de unión pero su escala se encuentra muy corrida hacía los números más pequeños), los resultados se encuentran resumidos en la tabla **??**

Motivo estructural	**Compuestos**	**Resultados positivos**
Urea	18883	12170
Carbamato	141	5
Acido Metoxiacético	4782	2146
Glicina	1926	714
Hidroxifuranona	164	0

Tabla 5.1: **Resultados de búsqueda de motivos estructurales en ZINC y de la búsqueda virtual en umaA**

Urea. Este grupo contiene 12170 resultados. La mayor parte de las drogas aquí encontradas

pueden realizar interacciones tanto con la cadena principal de la cisteína 34 como con la tirosina 231 y el glutámico 139. Por otra parte se observa una gran presencia interacciones de tipo *Pi-stacking* entre las drogas y la fenilalanina 198 y/o tirosina 262. Dada la gran cantidad de "positivos" se comentarán los primeros 10. En la figura **??** se presentan las estructuras de los mejores compuestos mientras que en la figura **??** se muestra en modo de unión de dos de ellos. Lo interesante de este conjunto de compuestos es que presentan una similitud química grande, salvo el compuesto 3, el resto pertenecen a una misma familia química, que se extiende más allá del grupo urea.

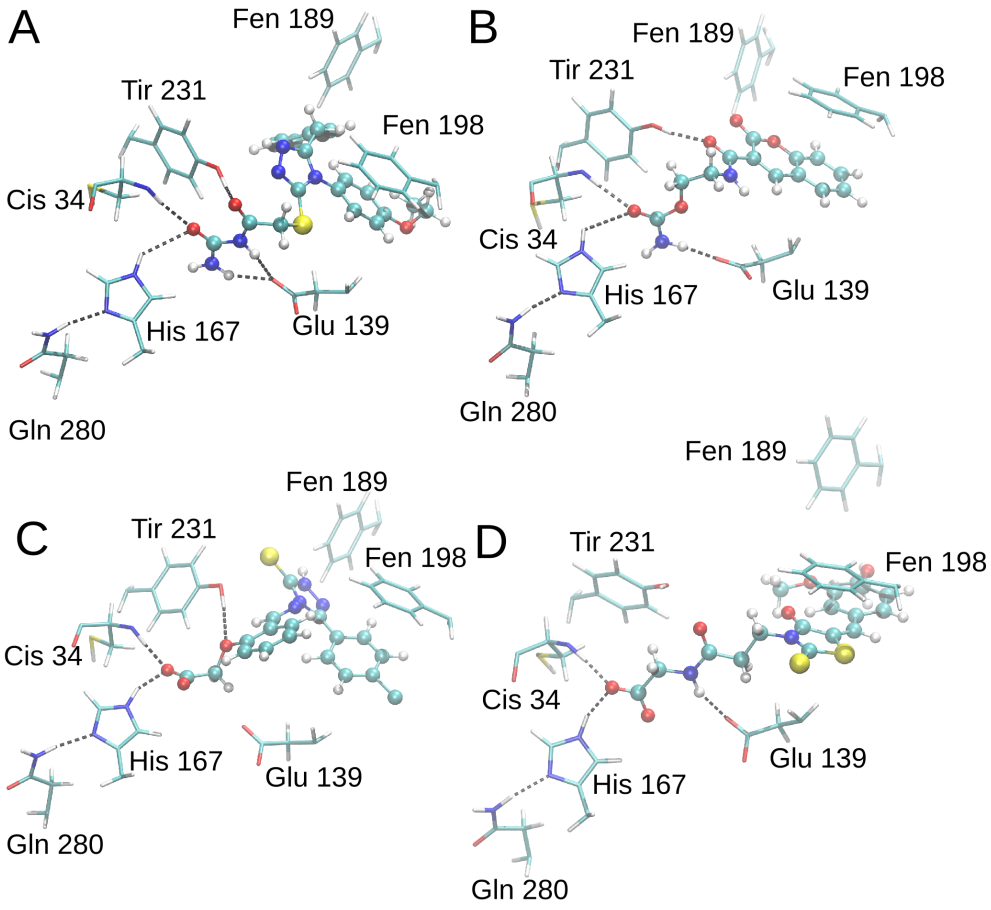


Figura 5.3: **Estructuras de *docking* sobre umaA de los distintos fragmentos** A) Urea (ZINC36859431), B) Carbamato (ZINC02024095), C) Ácido metoxiacético (ZINC06059535) y D) Glicina (ZINC05374970). Las lineas punteadas indican interacciones de tipo puente hidrógeno.

Carbamato. Este grupo contiene 141 compuestos que cumplen con el *restraint* farmacofórico. Solo 5 de ellos presentan un puntaje de interacción menor a -25 unidades, en la figura **??** se puede observar la estructura de los compuestos positivos. Al igual que con el motivo de Urea, los compuestos positivos presentan un *Pi-stacking* con fenilalanina 198, en algunos casos con la tirosina 230. A diferencia del motivo de Urea no se observa un motivo químico distintivo

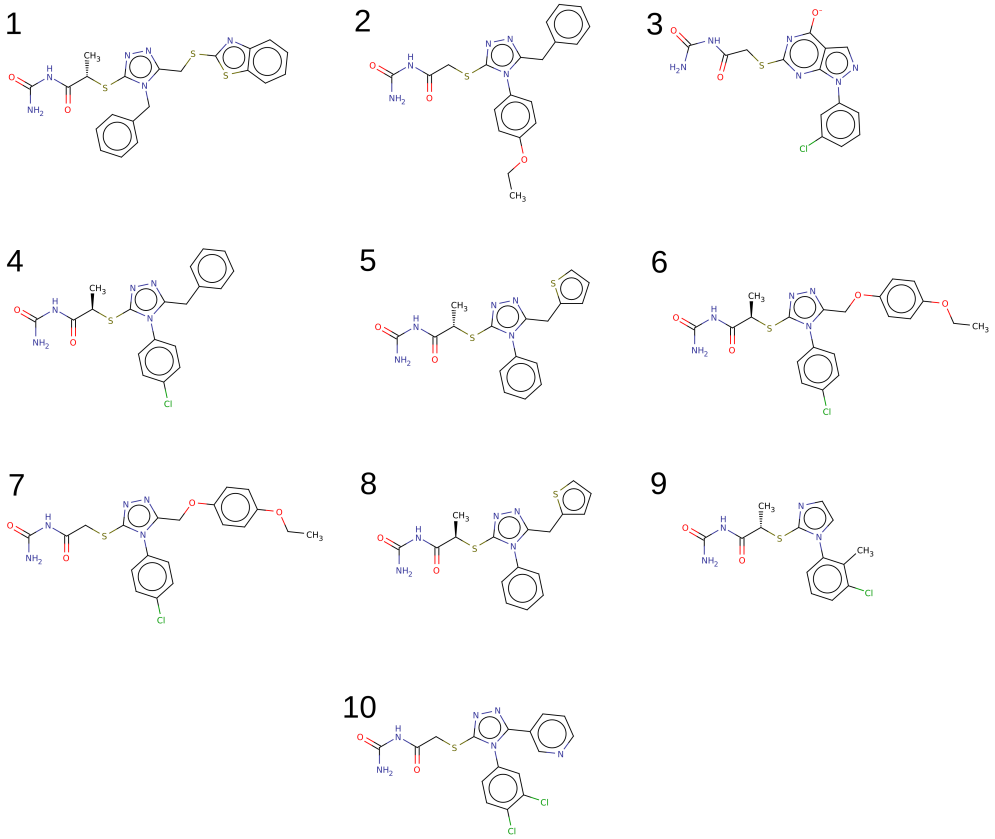


Figura 5.4: **Estructura de las mejores drogas basadas en el motivo de unión de Urea.** Los identificadores son los siguientes: 1)ZINC03506439 2)ZINC36859431 3)ZINC12740738 4)ZINC09570734 5)ZINC06993185 6)ZINC02621554 7)ZINC11022390 8)ZINC06993187 9)ZINC08710095 10)ZINC15274986

más allá del impuesto por el carbamato. En la figura **??**B se puede observar el modo de unión.

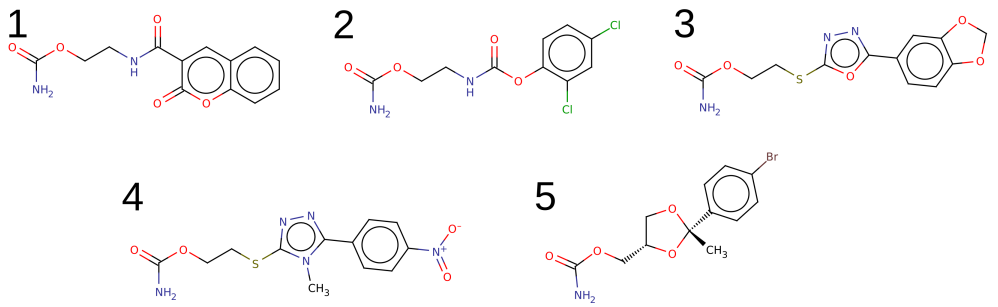


Figura 5.5: **Estructura de las mejores drogas basadas en el motivo de unión de Carbamato.** Los identificadores son los siguientes: 1) ZINC02024095 2) ZINC03158901 3) ZINC04891055 4)ZINC04966565 5)ZINC02008606

Ácido metoxiacético Este grupo presenta 2146 compuestos que cumplen con el *restraint* farmacofórico, 50 de ellos reúnen las características de unión como linea de corte, puntaje de interacción menor a -25 unidades. Comparten las interacciones descriptas para los fragmentos anteriores, *Pi-stacking* con fenilalanina 198 y tirosina 263. A diferencia del caso del motivo basado en urea no se observa una familia química claramente favorecida entre estos compuestos, más allá del evidente sesgo introducido al generar las búsquedas para construir la biblioteca.

Glicina Este conjunto presenta 714 compuestos que cumplen con el criterio de linea de corte (energía de interacción menor a -25 unidades). Además de las interacciones descriptas para los otros fragmentos (*Pi-stacking* con fenilalanina 198 y tirosina 230) Algunos compuestos presentan interacciones hidrofóbicas con las cadenas laterales de Leu 203 e Ile 193.

Hidroxifuranona En el caso de la hidroxifuranona, ningún compuesto logró cumplir con los *restraints* farmacofóricos, razón por la cual no se realizará ningún comentario sobre el modo de unión de los mismos.

5.3.4 Estudio del modo de unión por dinámica molecular y MM/PBSA

Para evaluar la estabilidad de las interacciones y la relevancia de las mismas para la unión a umaA (y sus similares, cmaA1-2, pcaA y mmaA2) se realizaron estudios de dinámica molecular y MM-PBSA de los mejores compuestos evaluados tanto desde el punto de vista de la función de puntuación como del modo de unión obtenido. Los números entre paréntesis detrás de cada compuesto indican cual de cual compuesto se trata en la figura respectiva de cada fragmento (Figura **??** para Urea, **??** para carbamato, **??** para metoxiacético y **??** para glicina).

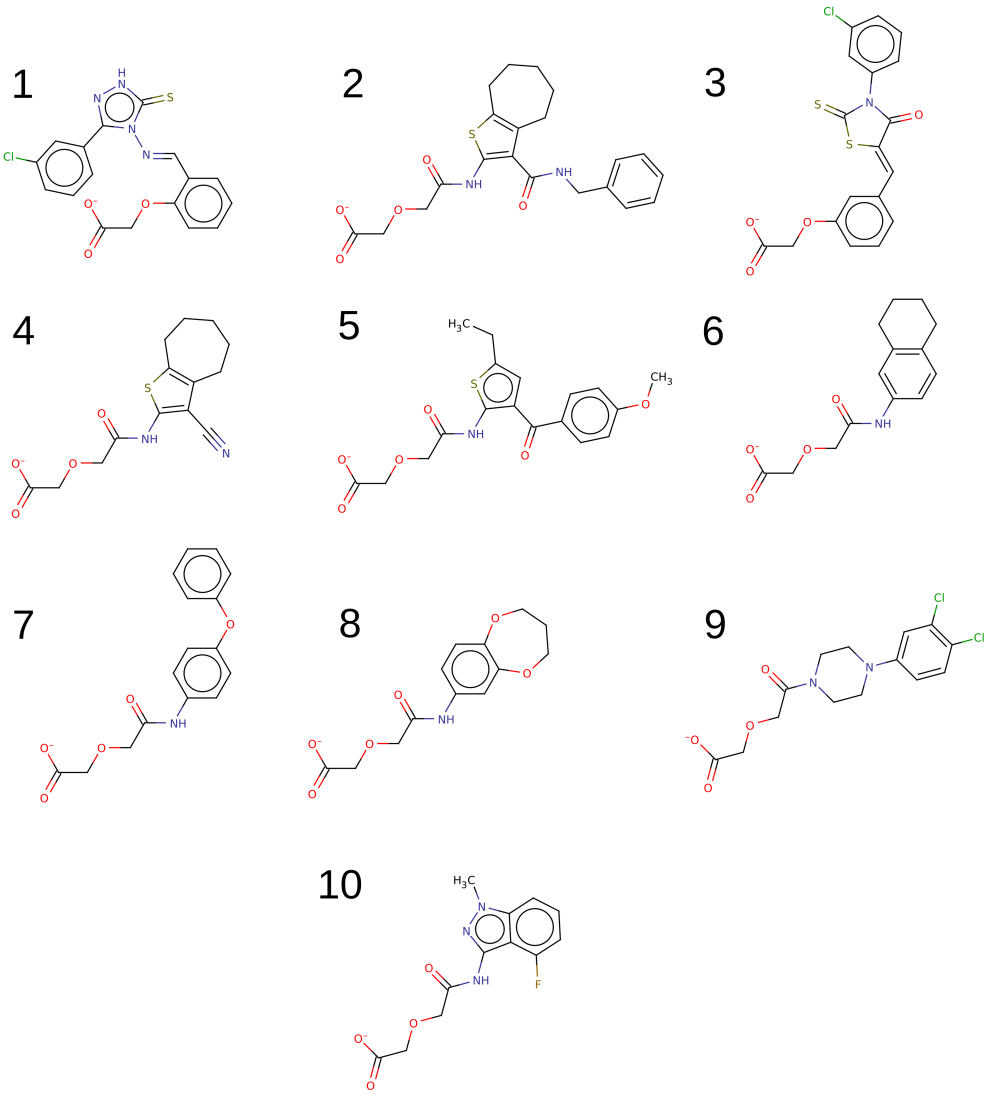


Figura 5.6: **Estructura de las mejores drogas basadas en el motivo de unión de ácido metoxiacético.** Los identificadores son los siguientes: 1)ZINC06059535 2)ZINC17090657 3)ZINC16399519 4)ZINC17031713 5)ZINC04334497 6)ZINC12375164 7)ZINC04616853 8)ZINC00117863 9)ZINC17072207 10)ZINC00118984

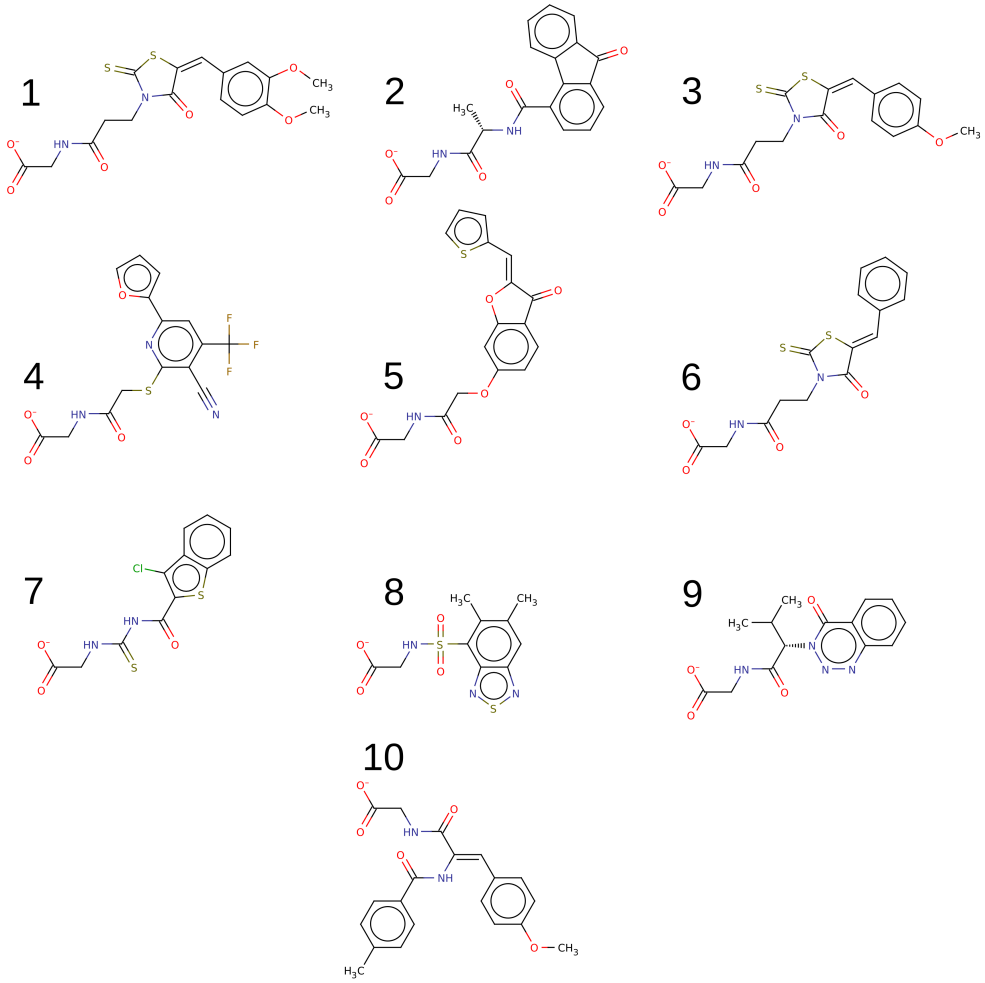


Figura 5.7: **Estructura de las mejores drogas basadas en el motivo de unión de Glicina.** Los identificadores son los siguientes: 1)ZINC05374970 2)ZINC05269172 3)ZINC05374959 4)ZINC08382611 5)ZINC06659663 6)ZINC35655933 7)ZINC02049261 8)ZINC00063553 9)ZINC02357332 10)ZINC16267236

Las drogas que se evaluaron fueron ZINC36859431 (2), ZINC12740738 (3), ZINC09570734, ZINC06993185 (5), ZINC02621554 (6) y ZINC11022390 (7) basadas en el fragmento de urea; ZINC02024095 (1), ZINC03158901 (2), ZINC04891055 (3) y ZINC04966565 (4) basadas en el fragmento de carbamato; ZINC06059535 (1), ZINC17090657 (2), ZINC16399519 (3), ZINC17031713 (4) y ZINC04334497 (5) basadas en el fragmento de metoxiacético; ZINC05374970 (1), ZINC05269172 (2), ZINC05374959 (3), ZINC08382611 (4) y ZINC06659663 (5) basadas en el fragmento de glicina. En total se corrieron 40ns de dinámica molecular clásica para cada uno de los distintos sistemas (20ns con *restraint* de distancia para mantener las interacciones y 20ns de dinámica libre). El esquema de trabajo seguido se puede observar en la figura **??**.

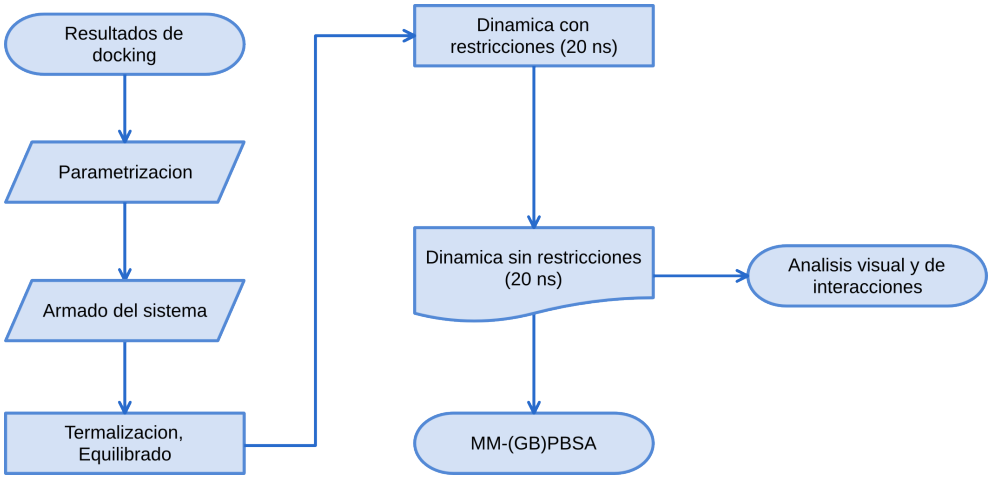


Figura 5.8: **Esquema de trabajo para el análisis de los mejores resultados de *docking* por dinámica molecular y MM-PBSA**

En primer lugar se analizó la estabilidad de los ligandos luego de 20ns de dinámica molecular, como así también cambios en los patrones de interacción respecto a los encontrados en el experimento de búsqueda virtual. En general las drogas basadas en urea intercambian el puente hidrógeno que formaban con cisteína 34 o con histidina 167 por la cadena lateral de treonina 33, que rota de su posición anterior cosa que le permite formar puentes hidrógeno actuando como donor con los ligandos. En general los ligandos permanecen bastante estables en el sitio activo de umaA (RMSD <2 Å). Salvo ZINC02024095 que cambie su set de interacciones, formando puente hidrógeno con Gln 97 con su cadena lateral, Ile 134 y Gli 70 con su cadena principal e interacciones Pi-Pi con fenilalanina 140 y ZINC09570734 que cambia sus interacciones por otras interacciones hidrofóbicas manteniendo las provenientes del farmacóforo. Ver figura **??**. A pesar de su salto en RMSD, se decidió analizar por MMPBSA estas drogas porque si bien el juego de interacciones cambia, se mantienen estables a lo largo de la dinámica.

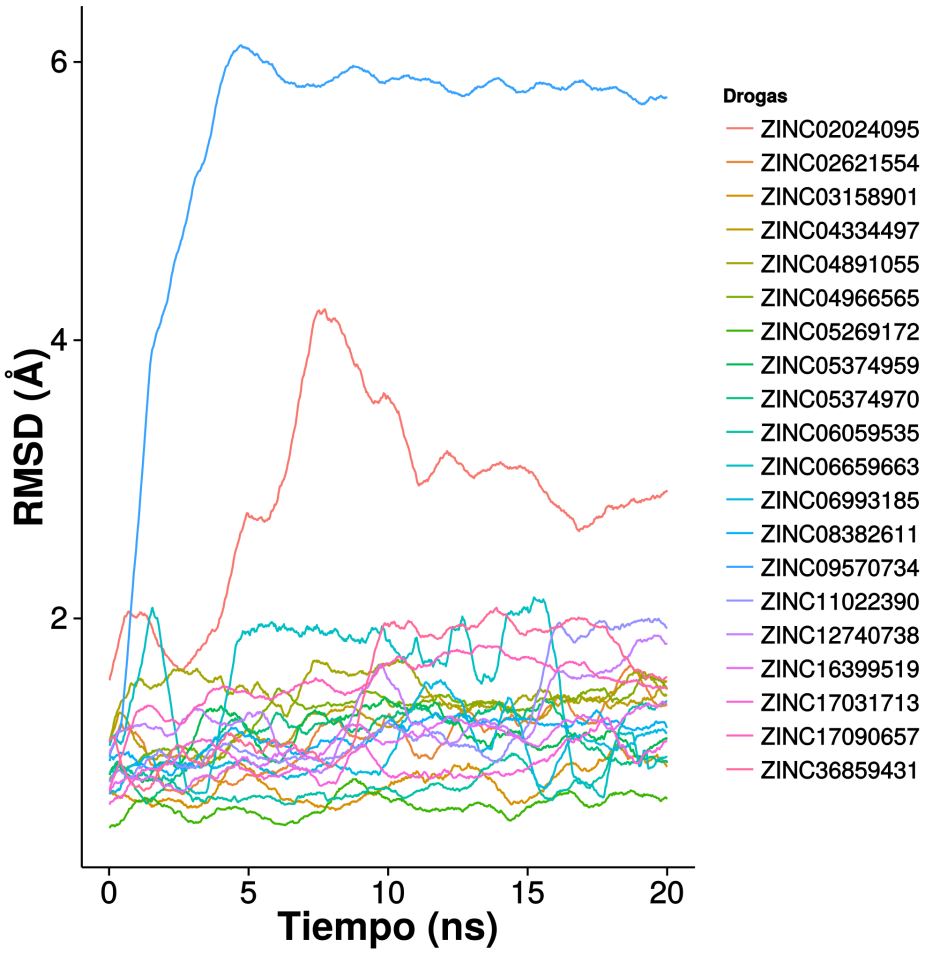


Figura 5.9: **Calculo de RMSD de las distintas drogas a lo largo de 20ns de dinámica molecular clásica**

Por otra parte se realizó un computo para estimar la energía libre de unión mediante la técnica de post-procesamiento MM-PBSA. Los resultados que se presentan en la tabla **??** se encuentran deconvolucionadas en una componente polar, formada por las interacciones electrostáticas y la componente polar de la energía libre de solvatación, y en una componente no polar, conformada por las interacciones de Van der Waals y la componente no polar de la solvatación estimada por Poisson-Boltzmann. Si se observa la tabla **??** se puede apreciar que el componente que define la interacción de todas las drogas aquí evaluadas es el No-Polar.

Compuesto	**Tipo**	**Polar**	**NoPolar**	ΔG
ZINC05374970	Glicina	13.46	-58.965	-45.506
ZINC11022390	Urea	26.218	-71.666	-45.448
ZINC05374959	Glicina	9.68	-54.775	-45.095
ZINC08382611	Glicina	10.363	-51.264	-40.901
ZINC06993185	Urea	15.998	-55.865	-39.868
ZINC02621554	Urea	31.023	-70.024	-39.001
ZINC04891055	Carbamato	9.961	-48.234	-38.273
ZINC09570734	Urea	18.771	-56.67	-37.9
ZINC06659663	Glicina	13.855	-50.348	-36.493
ZINC36859431	Urea	23.896	-58.139	-34.243
ZINC12740738	Urea	17.14	-50.848	-33.708
ZINC05269172	Glicina	21.858	-52.902	-31.044
ZINC16399519	Metoxiacético	25.78	-56.73	-30.95
ZINC17031713	Metoxiacético	13.254	-43.838	-30.585
ZINC04966565	Carbamato	16.168	-46.304	-30.137
ZINC06059535	Metoxiacético	25.843	-55.807	-29.963
ZINC03158901	Carbamato	16.264	-43.626	-27.363
ZINC04334497	Metoxiacético	25.098	-51.962	-26.864
ZINC17090657	Metoxiacético	33.492	-58.522	-25.029
ZINC02024095	Carbamato	19.26	-37.221	-17.96

Tabla 5.2: **Resultados numéricos de MMPBSA de los compuestos completos.** Se presenta una decomposición en la componente polar de la interacción (Electroestatática y Solvatación de PB) y la componente no polar de la interacción (VdW y la contribución no polar de la energía libre de solvatación). Las unidades de todos los datos reportados en esta tabla son kcal/mol.

Por otra parte, si se observa el resultado final, es decir el ΔG de unión estimado, la tabla se encuentra claramente enriquecida en compuestos con fragmentos de Glicina y Urea mientras que las compuestos basadas en ácido Metoxiacético y Carbamato parecen tener una mayor energía libre de unión. Dado que se desconoce si las compuestos basadas en glicina y urea tiene una energía libre de unión menor por sus fragmentos o por sus grupos R se decidió realizar nuevamente los cómputos de MM-PBSA en ausencia de los átomos del grupo R de cada com-

puesto. Esto permitirá comprender el origen de las diferencias entre los distintos fragmentos. Como MM-PBSA se trata de un esquema totalmente aditivo se puede estimar la contribución de cada uno de los grupos R en cada compuesto realizando la sustracción correspondiente. En las tablas **??** y **??** se presenta el resultado para los fragmentos y para los grupos R, respectivamente.

Compuesto	**Tipo**	**Polar**	**NoPolar**	ΔG	$\%\Delta G$
ZINC06993185	Urea	-8.63	-9.611	-18.241	45.75%
ZINC11022390	Urea	-1.035	-14.024	-15.059	33.13%
ZINC02621554	Urea	-1.719	-12.461	-14.18	36.36%
ZINC12740738	Urea	-3.688	-8.705	-12.394	36.77%
ZINC09570734	Urea	-1.866	-9.483	-11.35	29.95%
ZINC36859431	Urea	-2.848	-7.747	-10.595	30.94%
ZINC05374959	Glicina	-13.536	-7.139	-20.675	45.85%
ZINC05374970	Glicina	-8.692	-10.833	-19.525	42.91%
ZINC08382611	Glicina	-8.842	-8.537	-17.379	42.49%
ZINC06659663	Glicina	-8.72	-8.392	-17.112	46.89%
ZINC05269172	Glicina	-3.136	-9.811	-12.947	41.70%
ZINC06059535	Metoxiacético	-5.782	-10.697	-16.479	55.00%
ZINC17031713	Metoxiacético	-0.973	-9.849	-10.822	35.38%
ZINC17090657	Metoxiacético	-0.458	-7.737	-8.195	32.74%
ZINC16399519	Metoxiacético	3.571	-10.813	-7.242	23.40%
ZINC04334497	Metoxiacético	2.825	-7.943	-5.119	19.05%
ZINC04891055	Carbamato	-3.87	-9.378	-13.248	34.61%
ZINC04966565	Carbamato	0.322	-8.933	-8.611	28.57%
ZINC03158901	Carbamato	2.648	-10.429	-7.781	28.44%
ZINC02024095	Carbamato	4.565	-7.237	-2.672	14.88%

Tabla 5.3: **Decomposición de la energía libre de unión estimada por MMPBSA para los átomos que pertencen a los cuatro fragmentos descriptos (Glicina, Urea, Metoxiacético y Carbamato).** En la quinta columna se presentan los resultados de ΔG de unión para cada fragmento relativizado al total de ese compuesto

Al observar los resultados de la tabla **??** se puede concluir en primer lugar que ambas componentes, polar y no polar, son favorables para todos los casos de Urea y Glicina, siendo los más mejores nuevamente. Para el caso de Carbamato y metoxiacético la componente polar resulta desfavorable y la interacción esta gobernada por la componente no polar. Desde un punto de vista estructural esta diferencia entre el par Urea/Glicina y el par Carbamato/Metoxiacético se puede explicar por la posibilidad de realizar un puente hidrógeno fuerte con su respectivo grupo amida (HN-) con la cadena lateral de glutámico 139. Esta interacción no es posible con el ácido metoxiacético y resulta mucho más débil al realizarla con el grupo NH2 del Carbamato. Con lo cual de este análisis podemos concluir que el puente hidrógeno Glu 139-HN es importante

para el proceso de unión. Por otra parte, es interesante señalar que el componente no-polar es favorable para todos los fragmentos, probablemente dada la naturaleza cargada y polar de esta sección de la proteína, necesaria para la unión del ión bicarbonato. Es posible analizar la contribución porcentual de cada uno de los fragmentos a la energía libre de unión total de los compuestos. En el caso de los compuestos basados en Glicina, este fragmento puede explicar desde el 47% al 42% del ΔG de unión. En segundo lugar se encuentra los compuestos basados en Urea, con una dispersión un tanto mayor de 46% a un 30% del ΔG total. Salvo por el caso ZINC06059535, que presenta un porcentaje atípico de 55%, el resto de los compuestos basados en Metoxiacético se encuentran en valores entre 35% y 19% mientras que los basados en Carbamato se sitúan en valores entre 34% y 15%. Este resultado corrobora la tendencia observada en el caso del análisis con los compuestos completos, los fragmentos de Glicina y Urea tienen una mayor afinidad por el sitio de ión bicarbonato de umaA que Metoxiacético y carbamato.

Compuesto	**Tipo**	**Polar**	**NoPolar**	ΔG
ZINC06659663	Glicina	5.775	-43.419	-37.644
ZINC05374970	Glicina	27.253	-57.643	-30.389
ZINC04891055	Carbamato	20.637	-47.187	-26.55
ZINC05374959	Glicina	22.152	-48.133	-25.981
ZINC02621554	Urea	13.831	-38.856	-25.025
ZINC11022390	Urea	32.742	-57.563	-24.821
ZINC06993185	Urea	23.216	-47.636	-24.42
ZINC36859431	Urea	22.209	-45.917	-23.708
ZINC12740738	Urea	26.744	-50.393	-23.648
ZINC08382611	Glicina	19.205	-42.727	-23.522
ZINC04334497	Metoxiacético	22.273	-44.018	-21.745
ZINC06059535	Metoxiacético	24.628	-46.254	-21.626
ZINC17090657	Metoxiacético	15.846	-37.371	-21.526
ZINC09570734	Urea	20.828	-42.143	-21.314
ZINC04966565	Carbamato	14.227	-33.989	-19.763
ZINC16399519	Metoxiacético	13.616	-33.197	-19.581
ZINC05269172	Glicina	22.575	-41.956	-19.381
ZINC17031713	Metoxiacético	24.994	-43.091	-18.098
ZINC02024095	Carbamato	14.695	-29.984	-15.288
ZINC03158901	Carbamato	31.625	-45.11	-13.485

Tabla 5.4: **Descomposición de la energía libre de unión estimada por MMPBSA para los grupos R**. Todas las energías se encuentran reportadas en kcal/mol.

El análisis desarrollado arriba para los fragmentos se puede realizar para los grupos R de cada una de las compuestos, independientemente del fragmento al que pertenezcan. En la tabla

?? se pueden observar los distintos grupos R de las compuestos. En todos los casos el componente No-Polar gobierna la interacción, esto es esperable ya que el sitio de unión de las CMAS, y de umaA en particular, presenta una serie de aminoácidos hidrofóbicos y aromáticos que ayudan a estabilizar el estado de transición y la unión del lípido a las mismas. Los primeros 5 grupos R comparten una serie de propiedades interesantes, formados por lo menos por 2 anillos aromáticos con grupos sustituyentes que le brindan mayor polaridad. La importancia en la polaridad para la unión se puede evidenciar entre los compuestos ZINC05374970 y ZINC05374959, que pertenecen a la misma familia de compuestos, diferenciándose solamente en la presencia de un grupo eter adicional en ZINC05374970. Compuestos de mayor tamaño y menor polaridad como ZINC05269172 tienen una componente no-polar menos favorable. Compuestos demasiado polares, como ZINC08382611, no resultan muy favorables. Por otra parte los compuestos basados en glicina presentan interacciones con dos residuos hidrofóbicos como la Leu 203 y la Ile 193, no presentes en los otros fragmentos. Para el caso de las interacciones polares, existen un puente hidrógeno mediado por agua entre el grupo R y el la cadena lateral de Tyr 230 para ZINC06659663 mientras que para ZINC02621554 y ZINC04891055 este puente hidrógeno se presenta directamente entre el grupo R y Tyr 230. La presencia de algunas interacciones polares explica una parte de la mejora de la energía de unión de los grupos R arriba señalados respecto al resto, dado que presentan un set de interacciones hidrofóbicas similares entre ellos, ausentes en el resto de los compuestos.

Cabe destacar que los análisis aquí presentados se realizaron ignorando estas interacciones compuesto-agua-proteína.

En la figura **??** se pueden observar las interacciones que se encuentran conservadas en los distintos compuestos (marcadas con círculos rojos en la figura correspondiente). Muchas son de carácter hidrofóbico como la de Fenilalanila 198 o Leucina 285. Las interacciones polares más importantes son las ya señaladas, Cisteina 34, Treoina 33, Glutamico 139 y en menor medida Histidina 167 y tirosina 230. Cabe destacar que la estimación de energía aquí presentada se trata de una estimación hibrida dado que el componente de desolvatación en su formulación (Ver materiales y métodos) tiene una componente entrópica mientras que el resto de las componentes sólo son entálpicas.

Los compuestos que lucen más prometedores para la etapa de pruebas *in vitro* son los basa-

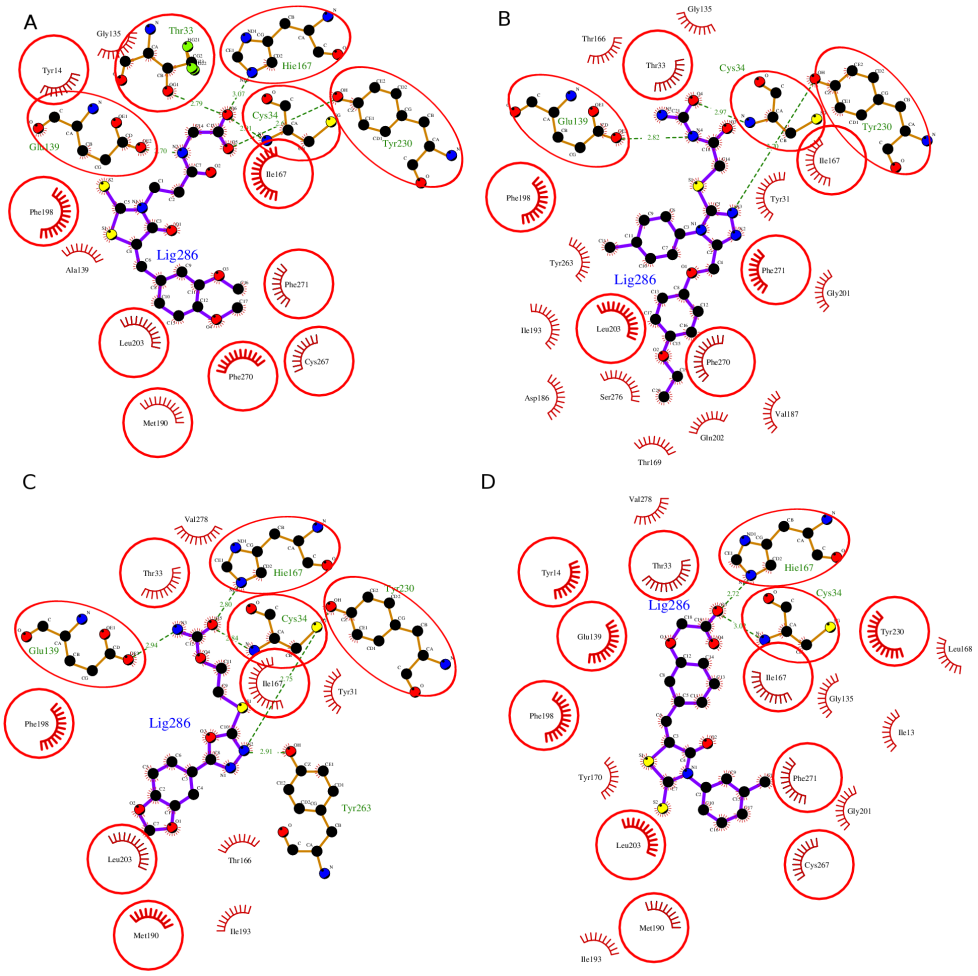


Figura 5.10: **Interacciones de la foto promedio de la dinámica molecular clásica de: A) ZINC05374970 B)ZINC11022390 C) ZINC04891055 D) ZINC16399519**

dos en urea y glicina. Presentan las mejores interacciones desde un punto de vista energético y estructural.

5.4 Discusión

En este capitulo se realizó una búsqueda virtual de posibles inhibidores para umaA, y en menor medida para todas las CMAS que comparten su estructura de sitio activo (cmaA1-2, pcaA y mmaA2) utilizando una característica puntual que comparte este grupo, la presencia de un ión bicarbonato en el sitio activo que resulta fundamental para la actividad catalítica de este grupo de enzimas. Solo cuatro de los cinco motivos de unión propuestos resultaron positivos del experimento de búsqueda virtual (Urea, Glicina, Carbamato y Ácido metoxiacético). Por el modo de unión encontrado mediante docking y dinámica molecular clásica estos compuestos evitarían la entrada del lípido a ser metilado y del ión bicarbonato.

Por otra parte, los compuestos basados en Urea y Glicina serian más prometedores según los resultados brindados por el análisis de MM-PBSA, dada la mayor fuerza de la interacción de puente hidrógeno que pueden realizar entre sus grupos amina y Glu 139. En segunda instancia, los grupos R presentes en algunas de los compuestos de la familia de Glicina, en particular, ZINC06659663, pueden realizar interacciones de puente hidrógeno mediados por aguas, esta capacidad le permite tener una mejor energía libre de unión de su grupo R. Es interesante destacar que no existen tendencias fuertes respecto de la energía libre de unión y la masa total de cada uno de los compuestos, esto indica que el componente polar también es necesario para la unión y no puede ser explicado únicamente por interacciones de tipo hidrofóbicas. En conclusión, los sustituyentes R deben contener alguno grupo polar capaz de aceptar puentes H de moléculas de agua para mediar interacciones entre ellas y umaA.

Si se realizara el experimento de docking en cmaA1, cmaA2 y pcaA que presentan el ión bicarbonato los compuestos basados en urea y glicina deberían funcionar. En la figura XXX se muestra una superposición de la estructura de umaA en complejo con ZINC06993185 con las tres CMAS detalladas arriba. Como es posible apreciar, muchas de las posiciones de las cadenas laterales se conservan en este subconjunto, principalmente las interacciones de unión a bicarbonato pero también las realizadas por los residuos aromáticos (Tyr y Phe principalmente) hecho que sugiere fuertemente que los compuestos descritos en este capítulo también podrían

funcionar como inhibidores de cmaA1, cmaA2 y pcaA.

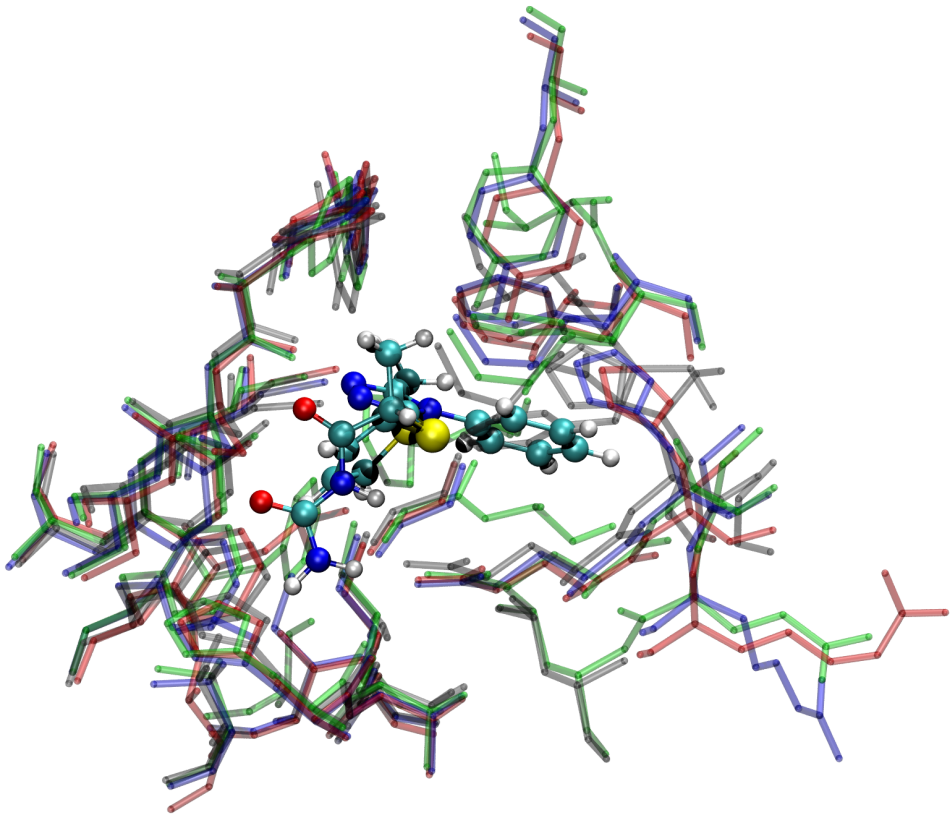


Figura 5.11: **Comparación estructural del sitio activo de umaA (verde), cmaA1 (azul), cmaA2 (rojo) y pcaA (gris) superpuesto con el resultado de docking de ZINC06993185 sobre umaA**

Existen drogas que trabajan a distintos niveles de la vía de síntesis del ácido micólico, como son etambutol o la isoniazida. Primeramente se había involucrado a las CMAS como el mecanismo de acción molecular de la droga thiacetazone dado que afecta la composición de los AM producidos. (Alahari et al., 2007) Sin embargo, luego se comprendió que los cambios asociados al uso de thiacetazone derivan de su efecto inhibitorio en las deshidratasas HadAB y HadBC (Coxon et al., 2013; Grzegorzewicz et al., 2012) evitando que se forme el sustrato para las CMAS y, por lo tanto, que no se produzcan modificaciones.

Por último, queda probar la efectividad de estos compuestos tanto *in vitro* y/o *in cellula* con los respectivos controles para probar el mecanismo molecular de acción de ser activos dado que es fundamental para validar los datos. Para ello se esta trabajando en la expresión y purificación de umaA como así también en su caracterización biofísica. En este punto, se tienen espectros de fluorescencia y de dicroísmo circular como controles de calidad de la expresión proteica. Por otra parte, dado que lo que se desea determinar es inhibición proteica es necesario contar

con un ensayo de actividad, o por lo menos de unión diferencial, que permita determinar si los compuestos se unen y, si se unen con una afinidad aceptable lo están haciendo al sitio correcto. Nos encontramos trabajando en un ensayo de FRET (Transferencia de energía resonante de Foster) entre los triptofanos de la proteína (presentes únicamente en el sitio activo de la misma) y ANS (una sonda fluorescente cuya fluorescencia cambia según la polaridad del entorno). Con dicho ensayo se podrá determinar una constante de afinidad para luego realizar experimentos de competencia entre esta sonda fluorescente y los compuestos mencionados en los resultados. Si bien sería deseable contar con un ensayo de actividad enzimática, el sistema presenta grandes dificultades para determinar la misma *in vitro*, principalmente por la disponibilidad de lípidos insaturados de cadena larga solubles.

Bibliografía

Alahari, Anuradha et al. (2007). “Thiacetazone, an antitubercular drug that inhibits cyclopropanation of cell wall mycolic acids in mycobacteria”. In: *PLoS One* 2.12, e1343.

Barkan, Daniel et al. (2012). “Mycobacterium tuberculosis lacking all mycolic acid cyclopropanation is viable but highly attenuated and hyperinflammatory in mice”. In: *Infection and immunity* 80.6, pp. 1958–1968.

Case, DA et al. (2014). “Amber 14”. In:

Coxon, Geoffrey D et al. (2013). “Synthesis, antitubercular activity and mechanism of resistance of highly effective thiacetazone analogues”. In: *PloS one* 8.1, e53162.

Grzegorzewicz, Anna E et al. (2012). “A common mechanism of inhibition of the Mycobacterium tuberculosis mycolic acid biosynthetic pathway by isoxyl and thiacetazone”. In: *Journal of Biological Chemistry* 287.46, pp. 38434–38441.

Hawkins, Gregory D, Christopher J Cramer, & Donald G Truhlar (1996). “Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium”. In: *The Journal of Physical Chemistry* 100.51, pp. 19824–19839.

Irwin, John J et al. (2012). “ZINC: a free tool to discover chemistry for biology”. In: *Journal of chemical information and modeling* 52.7, pp. 1757–1768.

Li, Li, Rong Chen, & Zhiping Weng (2003). “RDOCK: Refinement of rigid-body protein docking predictions”. In: *Proteins: Structure, Function, and Bioinformatics* 53.3, pp. 693–707.

Luo, Ray, Laurent David, & Michael K Gilson (2002). “Accelerated Poisson–Boltzmann calculations for static and dynamic systems”. In: *Journal of computational chemistry* 23.13, pp. 1244–1253.

Miyamoto, Shuichi, & Peter A Kollman (1992). “SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models”. In: *Journal of computational chemistry* 13.8, pp. 952–962.

OLBoyle, Noel M et al. (2011). “Open Babel: An open chemical toolbox”. In: *J Cheminf* 3, p. 33.

Onufriev, Alexey, Donald Bashford, & David A Case (2004). “Exploring protein native states and large-scale conformational changes with a modified generalized born model”. In: *Proteins: Structure, Function, and Bioinformatics* 55.2, pp. 383–394.

Ruiz-Carmona, Sergio et al. (2014). “rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids”. In: *PLoS Comput Biol* 10.4, e1003571.

Schrödinger, LLC (2011). “LigPrep, version 2.5”. In: *New York, NY*.

Capítulo 6

Conclusiones generales y perspectivas.

6.1 Conclusiones

Como bien señalamos en la introducción de esta tesis, la búsqueda de fármacos es una área de activo desarrollo en donde la bioinformática puede realizar contribuciones significativas al aportar valiosas herramientas a la hora de proponer blancos moleculares y compuestos candidatos para modular su actividad. En el caso particular de *Mycobacterium tuberculosis* existe una creciente necesidad de encontrar nuevos compuestos con actividad bactericida para la fase de latencia, para la cual aún no existen fármacos disponibles. Por esta razón plateamos tres objetivos para ayudar en la búsqueda de nuevos fármacos para combatir la Tuberculosis. El primero es la elección de nuevos blancos moleculares, etapa crítica de cualquier proyecto de desarrollo farmacéutico, con tal finalidad fue creada TuberQ. El segundo y tercer objetivo, en realidad se encuentran combinados dado que comprender el mecanismo de reacción contribuye a generar hipótesis sobre cual sería el mecanismo optimo de inhibición para un grupo de enzimas. En resumen hemos partido de un genoma, elegido nuevos blancos moleculares y propuesto nuevas moléculas para inhibir el crecimiento de *Mtb*.

Fruto del primer objetivo hemos desarrollado la herramienta denominada TuberQ que es capaz de facilitar la búsqueda de nuevos blancos principalmente en base a criterios de drogabilidad contextual. Esta aplicación web permite a cualquier investigador del área de la Tuberculosis tener en una base de datos centralizada la información acerca de la *bindability*, que residuos forman parte del sitio activo, la sensibilidad a estrés de ERON, la información de expresión en condiciones que imitan la infección y la importancia dentro de la red metabólica del bacilo.

Dicha información se encuentra disponible de forma pública, e incluso, tanto los modelos como la información de *bindability* es posible bajarlos para que cada usuario pueda realizar su propio análisis. Justamente, una de ventajas de TuberQ es que no impone ningún modelo de análisis de los datos a los usuarios, cada uno puede analizar los blancos con los criterios que elija y pesarlos de la forma que le parezca más relevante para su proyecto. Esto es un punto de diferencia con muchos estudios realizados en donde los resultados son presentados de forma cerrada en una tabla que indica cuáles son los mejores blancos para tratar la Tuberculosis (o cualquier patología), como los desarrollados con el grupo de (Raman, Yeturu, & Chandra, 2008; Anand, & Chandra, 2014). Adicionalmente, el usuario puede (previo pedido a los autores) subir su propio conjunto de datos para utilizarlos como otro criterio más a la hora de construir su propia función de puntuación. Por otra parte, si bien TuberQ fue pensado particularmente para trabajar con el genoma de *Mtb.* y sus problemas biológicos inherentes, la herramienta y los *pipelines* bioinformáticos desarrollados se pueden extender a cualquier genoma de interés. Por ejemplo, nuestro grupo ha utilizado la herramienta para producir una anotación estructural y predicción de blancos en *Corynebacterium pseudotuberculosis* (Radusky et al., 2015) y se encuentra en desarrollo la aplicación a *Klebsiella pneumoniae*.

En relación al segundo y tercer objetivo, la búsqueda virtual y el estudio computacional de la familia de las *Cyclopropane Mycolic Acid Synthases* resultan complementarios. Conocer el mecanismo de reacción de las CMAS permitió elegir motivo particular de unión de un subconjunto de ellas (las que utilizan bicarbonato para su reacción) como base para buscar compuestos que imiten el modo de unión. Además se evidenciaron otros elementos relevantes para que la reacción enzimática se produzca, como un residuo de ácido glutámico o la presencia de tirosinas en el sitio activo. Dicha información no fue utilizada directamente a la hora de realizar la búsqueda virtual pero, a la hora de analizar los compuestos encontrados mediante *docking* y dinámica molecular, son aquellos que forman interacciones con el ácido glutámico y la tirosina los que presentan la mejor energía de unión predicha. A pesar de no haber sido desarrollado en principio con este objetivo, dadas las similitudes existentes entre las CMAS, es posible que los compuestos encontrados funcionen en toda la familia en un rango similar de concentraciones, es decir que funcionen como inhibidores polifarmacológicos. Esto es de particular interés porque los cambios en el fenotipo de infección se observan al realizar mutaciones en un grupo impor-

tante de la familia de las CMAS y no en forma individual(Barkan et al., 2012). Por esta razón resulta importante que los compuestos posean acción polifarmacológica dentro de la familia de las CMAS.

6.2 Perspectivas

Como en todo trabajo científico, han quedado cosas pendientes que debido al tiempo acotado del trabajo no se han podido desarrollar. En general todos los puntos pendientes que quedan de esta tesis persiguen la validación experimental de los resultados computacionales aquí presentados. Particularmente realizar experimentos de mutagénesis condicional dirigida contra los genes miembros de la vía de síntesis de micotiol (ino1) o de lipoato (lipA y lipB) validaría el *pipeline* de búsqueda de blancos moleculares aquí presentado. Por otra parte, los resultados de la búsqueda virtual y mecanismos de reacción de CMAS requiere de un enfoque mixto bioquímico y microbiológico. Desde el punto de vista bioquímico, como prueba de concepto, se dispone de la enzima umaA recombinante caracterizada por métodos biofísicos (Fluorescencia de triptofanos, dicroísmo circular en el UV cercano y lejano). Esta caracterización permitió realizar ensayos, por ahora en estado preliminar y por eso no incluidos en esta tesis, de unión de ANS a umaA. El ANS (*8-Anilinonaphthalene-1-sulfonic acid*) es una sonda hidrofóbica que presenta un aumento significativo del rendimiento cuántico de emisión de fluorescencia cuando se encuentra en entornos hidrofóbicos. Se planea utilizar esta capacidad para realizar un ensayo de FRET entre los triptofanos de la proteína (recordar que existe uno en el sitio de unión de los compuestos según el *docking*) y la sonda ANS (para tener una referencia del sitio activo de umaA se puede observar la figura **??**). Una vez probado esto, se realizarán los experimentos de competencia de los compuestos por el sitio de ANS.

Una vez realizada la validación experimental, sería necesario optimizar los compuestos teniendo en cuenta las propiedades particulares de la membrana de *Mtb.*, para facilitar su entrada a la bacteria (Dartois, & Barry, 2013), aspecto que ha sido ignorado en la primera aproximación aquí presentada. Por otra parte, también es necesario desarrollar modificaciones que mejoren el modo de unión con mmaA4 dado que cumple un rol fundamental en la modulación de la respuesta inmune del hospedador siendo mucho más efectivos los compuestos tanto desde un punto de vista del modo de acción como a la posible aparición de mutaciones que generen

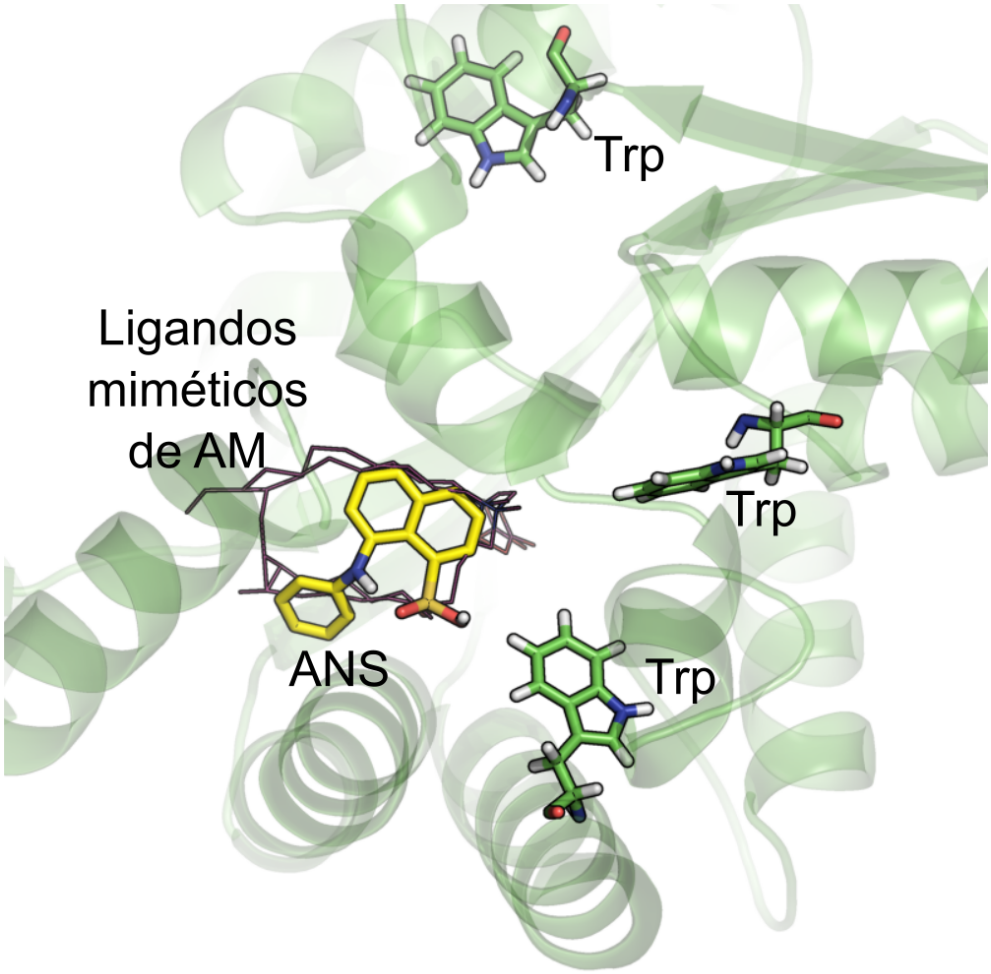


Figura 6.1: **Sitio activo de umaA**. En la figura se muestra el sitio activo de umaA con todos los triptofanos de la proteína (en verde) y la mejor estructura de *docking* (realizado con rDock) de ANS (*8-Anilinonaphthalene-1-sulfonic acid*, amarillo) y análogos de ácidos micólicos (en violeta).

resistencia si la droga actúa en un número mayor de blancos.

Existen otra serie de experimentos, más que nada microbiológicos, que permitirían responder preguntas básicas del funcionamiento de las CMAS, dado que actualmente no es posible medir la actividad enzimática de estas proteínas *in vitro*. Principalmente preguntas relacionadas con la interacción proteína-proteína de las CMAS con los miembros del complejo FAS II(InhA, hadAB/BC), cuyo rol en el proceso de síntesis de ácido micólico es desconocido. Tampoco se comprenden los determinantes moleculares entre la selectividad de las CMAS por el sitio distal y proximal de olifinas del ácido micólico inmaduro, aunque dicha selectividad podría estar modulada con la proteína transportadora de acilos (ACP), responsable de mediar la interacción de los AM de cadena larga y las proteínas encargadas de la síntesis y modificación del mismo. Experimentos de interacción proteína-proteína *in cellula* realizados mediante la técnica FRET

podrían ayudar a dilucidar si estas interacciones suceden dentro del bacilo y cuál es su en la síntesis de ácido micólico.

En resumen, en este trabajo de tesis hemos desarrollado técnicas de selección de blancos moleculares utilizando información genómica y de expresión combinándola con diferentes técnicas bioinformáticas para obtener una función propia de drogabilidad adaptada a la biología particular de *Mtb.* pero cuya aplicación no está limitada a este organismo. Además hemos propuesto, mediante técnicas de búsqueda virtual y dinámica molecular, compuestos de 4 familias distintas que podrían tener efecto bactericida/bacteriostatico en *Mycobacterium tuberculosis.*

Bibliografía

Anand, Praveen, & Nagasuma Chandra (2014). “Characterizing the pocketome of Mycobacterium tuberculosis and application in rationalizing polypharmacological target selection”. In: *Scientific reports* 4.

Barkan, Daniel et al. (2012). “Mycobacterium tuberculosis lacking all mycolic acid cyclopropanation is viable but highly attenuated and hyperinflammatory in mice”. In: *Infection and immunity* 80.6, pp. 1958–1968.

Dartois, Véronique, & Clifton E Barry (2013). “A medicinal chemists’ guide to the unique difficulties of lead optimization for tuberculosis”. In: *Bioorganic & medicinal chemistry letters* 23.17, pp. 4741–4750.

Radusky, Leandro G et al. (2015). “An integrated structural proteomics approach along the druggable genome of Corynebacterium pseudotuberculosis species for putative druggable targets”. In: *BMC Genomics* 16.Suppl 5, S9.

Raman, Karthik, Kalidas Yeturu, & Nagasuma Chandra (2008). “targetTB: a target identification pipeline for Mycobacterium tuberculosis through an interactome, reactome and genome-scale structural analysis”. In: *BMC systems biology* 2.1, p. 109.

Anexos

Publicaciones

.1 Publicaciones que surgieron de resultados de este trabajo de Tesis Doctoral.

- **Lucas Alfredo Defelipe**; Dario Fernandez Do Porto; Pablo Ivan Pereira Ramos; Marisa Fabiana Nicolas; Ezequiel Sosa; Leandro Radusky; Esteban Lazarotti; Adrian Gustavo Turjanski; Marcelo Adrián Martí. A Whole genome bioinformatic approach to determine potential latent phase specific targets in Mycobacterium tuberculosis. Tuberculosis (Edinb).Filadelfia: CHURCHILL LIVINGSTONE. 2015 vol. n. p -

- Radusky, Leandro G*; **Defelipe, Lucas Alfredo*** ;Lanzarotti, Esteban; Luque, Javier; Barril, Xavier; Marti, Marcelo Adrián; Turjanski, Adrian Gustavo. TuberQ: a Mycobacterium tuberculosis protein druggability database.. Database The Journal of Biological Databases and Curation.: Oxford University Press. 2014 vol. n. p1 - 10. issn 1758-0463. *Ambos autores contribuyeron equitativamente

Existe una publicación en preparación que contiene los resultados presentados en el capítulo 4.

.2 Publicaciones no relacionadas directamente con este trabajo de Tesis Doctoral.

- Cardama, Ga; Comin, J; Hornos, L; Gonzalez, N; **Defelipe, Lucas Alfredo**; Turjanski, Adrian Gustavo; Alonso, Df; Gomez, De; Lorenzano Menna, P. Preclinical development of novel Rac1-GEF signaling inhibitors using a rational design approach in highly aggressive breast cancer cell lines. ANTI-CANCER AGENTS IN MEDICINAL CHEMISTRY.:

BENTHAM SCIENCE PUBL LTD. 2013 vol. nº. p - . issn 1871-5206.

- Dumas, Victoria Gisel; **Defelipe, Lucas Alfredo**; Petruk, Ariel Alcides; Turjanski, Adrian Gustavo; Marti, Marcelo Adrián. QM/MM study of the C—C coupling reaction mechanism of CYP121, an essential Cytochrome p450 of *Mycobacterium tuberculosis*. PROTEINS: STRUCTURE, FUNCTION AND GENETICS.: WILEY-LISS, DIV JOHN WILEY & SONS INC. 2013 vol. nº. p - . issn 0887-3585.

- Mayoral, J.G.; Leonard, K.T; Nouzova, M.; **Defelipe, Lucas Alfredo**; Turjanski, Adrian Gustavo; Noriega, F.G.. Functional Analysis Of A Mosquito Short-Chain Dehydrogenase Cluster. ARCHIVES OF INSECT BIOCHEMISTRY AND PHYSIOLOGY.: WILEY-LISS, DIV JOHN WILEY & SONS INC. 2013 vol.82 nº2. p96 - 115. issn 0739-4462.

- Petruk, Ariel Alcides*; **Defelipe, Lucas Alfredo***; Rodriguez Limardo, Ramiro Gonzalo; Bucci, Hernan; Marti, Marcelo Adrián; Turjanski, Adrian Gustavo. Molecular dynamics simulations provide atomistic insight into hydrogen exchange mass spectrometry experiments. JOURNAL OF CHEMICAL THEORY AND COMPUTATION.Washington: AMER CHEMICAL SOC. 2013 vol.9 nº1. p658 - 669. issn 1549-9618. *Ambos autores contribuyeron equitativamente.

- Baquedano, Sonia M; Ciaccio M; Marino R; Perez Garrido N; Ramirez P; Maceiras M; Turjanski, Adrian Gustavo; **Defelipe, Lucas Alfredo**; Rivarola Ma; Belgorosky A.. A A novel missense mutation in the hsd3b2 gene, underlying nonsalt-wasting congenital adrenal hyperplasia. New insight into the structure-function relationships of 3-Beta-hydroxysteroid dehidrogenase type II. JOURNAL OF CLINICAL ENDOCRINOLOGY AND METABOLISM.: ENDOCRINE SOC. 2014 vol. nº. p - . issn 0021-972X.

- Zeida, Ari; Guardia, Carlos M; Lichtig, Pablo; Perissinotti, Laura L. ; **Defelipe, Lucas Alfredo**; Turjanski, Adrian Gustavo; Radi, Rafael; Trujillo, Madia; Estrin, Dario. Thiol redox biochemistry: insights from computer simulations. Biophysical Reviews.: Springer Berlin Heidelberg. 2014 vol.6 nº1. p27 - 46. issn 1867-2450.

- **Defelipe, Lucas Alfredo**; Lanzarotti, Esteban; Gauto, Diego; Marti, Marcelo Adrián; Turjanski, Adrian Gustavo. Protein Topology Determines Cysteine Oxidation Fate: The

Case of Sulfenyl Amide Formation among Protein Families. PLOS COMPUTATIONAL BIOLOGY.San Francisco: PUBLIC LIBRARY SCIENCE. 2015 vol.11 n3. p - . issn 1553-734X.

- Chaves, Alejandro; Eberle, Silvia Eandi; **Defelipe, Lucas**; Pepe, Carolina; Milanesio, Berenice; Aguirre, Fernando; Fernandez, Diego; Turjanski, Adrian; Feliú-Torres, Aurora; Two novel DNA variants associated with glucose-6-phosphate dehydrogenase deficiency found in Argentine pediatric patients,Clinical Biochemistry 2016,Elsevier. En prensa

Agradecimientos

Resulta muy complicado resumir en pocas palabras de agradecimiento a todas las personas que ayudaron a la realización de este trabajo. Muchas personas han contribuido indirectamente desde un punto de vista científico y/o humano para que esta tesis esté escrita y tal vez sus nombres no se encuentren plasmados mas que nada por un olvido del autor y no porque su rol no haya sido importante en ese momento.

A mis dos directores Adrián T. y Marcelo M., totalmente complementarios (aunque mucha gente no puede llegar a comprender eso) por confiar en mi para encarar un proyecto en sociedad. Por la infinita paciencia que me tuvieron. Por las largas discusiones (científicas y no). Y más que nada, por hacer del grupo un lugar agradable para trabajar contagiando ambos su entusiasmo por hacer ciencia.

A Javier Santos, por abrirme la puertas de su laboratorio experimental para realizar una parte importante del trabajo experimental de mi tesis, que si bien no está plasmado en este documento tomó dos años poner a punto. A pesar de que esta figura no exista, fue mi tercer director prestándome de su tiempo y su mente al pensar experimentos juntos.

Al grupo de los Prof. Javier Luque y Xavier Barril con los cuales discutimos desde el comienzo el *pipeline* de selección de blancos, siendo de gran ayuda sus comentarios y con quienes aprendí a realizar *Virtual Screening*.

A Ramiro Rodríguez Limardo, por ser la persona que me enseño durante mis primeros años de doctorado todo lo referente a la simulación computacional. Es muy probable que sin su ayuda hubiera sido mucho más difícil obtener los resultados presentados en esta tesis.

A Esteban Mocskos, por ser mi referente, y salvaquilombos, en los quehaceres informáticos, siempre dispuesto a ayudar.

A Raúl Esteban Ithuralde, por las discusiones políticas (universitarias, nacional, etc.) que hemos tenido a lo largo de estos años, han aportado mucho a mi forma de ver el mundo.

A Gonzalo Parra, por ser un gran amigo y colega, por charlas infinitas de lo que significa para nosotros hacer ciencia y de la vida.

Federico Osman, por prestar tu invaluable ayuda en el capítulo de las CMAS.

A todos mis compañeros de QB6, QB10, QB65, E1 y el Labo 8 y de otros labos. Algunos se fueron, otros entraron hace relativamente poco pero la ciencia es una actividad colectiva, sin estas personas buena parte del trabajo que realizo a diario sería mucho mas duro. Lanza, Lean R., Mode, Juan Angio, Hernan B., Martin D., Elias, Dipa, Clau, Vicky, Lu, Ari, Petruk, Romerito, Nacho B., Pablo L., Juan Pablo A., Juan Pablo B., Sol, Osvaldo, Marianito, Fer B., Diego G., Diego H.,Charly, Fede, Rodri, Nico F., Uriel, Ernesto, Juan R., Diego V., Nano, Ale F., Will, Martin N., Wanda, Santi F. y Nacho C, Migue y Sebas.

A mis compañeros de agrupación, Sumatoria, que me han permitido desarrollarme en el mundo de la política universitaria permitiéndome participar siendo ”firmero” y luego consejero brindándome otra perspectiva de como las cosas suceden.

A Agus, Flor, Andrés, Pau, Pauli, Guarra, Ale, Jere, Maru, Vicky, Ceci y el Chino, mis amigos de la facultad, que transitaron conmigo la carrera de grado y, muchos de ellos, están en el mismo camino tortuoso de realizar un doctorado con los cuales hemos compartido cientos de horas de catarsis y de celebraciones y, probablemente, seguiré compartiendo.

A Rodri, Mati, Maxi, Lucho y Emi, mis amigos de la vida. Por bancarme todos estos años mis locuras.

A todo el personal de las Facultades de Ciencias Exactas y Naturales y de Farmacia y Bioquímica de la Universidad de Buenos Aires, cuya labor puede resultar invisible y solo se ”nota” en su ausencia pero que resulta fundamental para que todos los que habitamos estos lugares podamos realizar con normalidad nuestro trabajo.

Al Estado nacional por financiar mis estudios universitarios de grado (a través de la gratuidad) y de posgrado (con una beca CONICET).

Al pueblo argentino por haber aportado los recursos financieros necesarios para la realización de este trabajo a través de sus impuestos.

A mis padres Alfredo y Alicia y mis hermanas Ana y Juliana, por ser un apoyo emocional incondicional durante todos estos años.

A mi compañera de vida, Mariana, que hace relativamente poco que nos conocemos pero

que ha sido crucial para la última etapa de esta tesis dándome soporte moral, tolerando largas jornadas laborales, incluso los fines de semana brindándome su cariño y paciencia.

Figuras

1.1 **Número de pares de bases (en miles de millones) depositados en el NCBI** (National Center for Biotecnology Information), NIH, USA, en azul. En rojo se indica el numero de usuarios por día. Las flechas indican la aparición de herramientas o bases de datos a lo largo de los años, como BLAST o OMIM. Gráfico del National Library of Medicine. 3

1.2 **Numero acumulado total de estructuras depositadas en el PDB por año.** Elaborado en base a datos del RSCB 4

1.3 **Procedimiento de modelado por homología.** El modelado requiere de una estructura que actúa como molde y la secuencia de ambas, que debe ser alineada tomando en cuenta la presencia de información estructural en el molde. El modelado se realiza por partes, tomando en primer lugar las restricciones derivadas de posicionar la cadena principal (representada como los carbonos α), luego se posiciona la cadena principal de regiones de las cuales no se dispone de información, en general flexibles, mediante técnicas *ab initio*. Por último, se insertan las cadenas laterales y se realiza una optimización global del modelo. 6

1.4 **Esquema de trabajo para el desarrollo de fármacos.** El esquema se puede dividir en dos etapas, una pre-clínica y una clínica. En la primer etapa se realizan todos los ensayos de búsqueda, validación, optimización y seguridad de las moléculas encontradas. En la segunda etapa se realizan los ensayos clínicos de efectividad y seguridad con poblaciones cada vez más grandes y, una vez aprobado, se pasa a la etapa de farmacovigilancia. Tomado de (Roses, 2008) . . 9

1.5 **Cantidad de muertes de TB a nivel global en cientos de miles.** Tomado de (WHO, 2014) 14

1.6 **Porcentaje de casos de MDR TB sobre el total de casos.** Tomado de (WHO, 2014) 14

1.7 **Representación esquemática de la pared de *Mycobacterium tuberculosis*** La pared se encuentra compuesta por tres componentes covalentemente unidos, el ácido micólico (en verde), los peptidoglicanos (en gris) y los arabinogalactanos (en celeste). La parte externa esta compuesta por la capsula (polisacaridos, glucano y arabinomanan) y lípidos libres complejos que se intercalan entre los AM. Adaptada de (Abdallah et al., 2007) 17

1.8 **Vía de síntesis del ácido micólico y su exportación a la pared**. En celeste están marcadas las proteínas involucradas en cada reacción. Las lineas rojas indican inhibición de ese paso por una de las drogas señaladas. Los asteriscos (*) indican que la proteína es fosforilada. Adaptado de (Marrakchi, Lanéelle, & Daffé, 2014) 18

1.9 **Proceso de infección de *Mtb.*.** Una vez en los pulmones, cuyo ingreso se originó por aerosoles, el bacilo es fagocitado por un macrófago alveolar que induce una respuesta proinflamatoria. Este tipo de respuesta es la que define a la enfermedad, formando un tubérculo o granuloma, en donde los macrófagos infectados son rodeados por células espumosas gigantes (células derivadas de macrófagos fusionados, multinucleadas y con gotas lipídicas) y un manto exterior de linfocitos. En esta etapa la infección no se propaga. Ante alguna falla del sistema inmune, esta barrera de contención de rompe y se produce la infección. Tomado de (Russell, 2001) 20

2.1 **Búsqueda de soluciones de *docking* mediante algoritmos genéticos lamarkianos.** A la izquierda se muestra un esquema del ”cromosoma” y la estructura de datos que representa junto con las variables a optimizar. A la derecha se muestra un gráfico de como es esa búsqueda en el espacio de variables (eje X) en relación con la función de puntuación (eje Y). Esquema de búsqueda tomado de el manual de Autodock 4. 47

3.1 **Representación esquemática del *pipeline* de TuberQ.** A partir del genoma traducido (los ORFs) y utilizando el PDB, Modeller, BLAST ,HMMer y PFAM es posible determinar si existen estructuras o si es necesario modelarlas además de determinar el/los dominios que le corresponden a cada ORF. Luego se pueden realizar sobre el estructuroma calculos adicionales como la *bindability* mediante fpocket, la determinación de sitios activos o la sensibilidad a ERON por presencia de Cisteinas y/o Tirosinas. 55

3.2 **Distribución de bolsillos proteicos en *Mtb.* de acuerdo a su clasificación por DS**. Clasificamos a los bolsillos en cuatro categorías distintas: No drogables si su puntaje se encuentra entre 0 y 0.2, probablemente drogable entre 0.2 y 0.4, drogable si se encuentra entre 0.4 y 0.7 y altamente drogable si se encuentra entre 0.7 y 1. 59

3.3 **Representación de los resultados de la búsqueda.** Cada triada UniProt-PFAM-Estructura representa una entrada diferente en la base de datos. Se puede elegir agrupar las entradas de UniProt al tildar 'Group by UniProtID' 63

3.4 **Solapa *Summary*.** En la solapa *Summary* se encuentra la información de asignación a UniProt, PFAM y PDB y los correspondientes links hacia esas bases de datos; la información acerca de la asignación de dominio y determinación de estructura realizados por el programa HMMer y BLAST respectivamente y el alineamiento entre la proteína de *Mtb* y su homologo más cercano en el genoma humano. 64

3.5 **Solapa *Structure*.** En en el panel superior se muestra la información sobre la drogabilidad máxima, la presencia de drogas o los residuos reportados en CSA. En el panel central se encuentra el visualizador de estructuras. El panel derecho permite al usuario controlar el visualizador (decidir que mostrar y como mostrarlo). Debajo se presenta, mayor información sobre cada entrada, como es la cantidad total de bolsillos encontrados, los residuos con desvíos en su pKa en solución (realizado con el programa propKa) o la presencia de metales. . . . 65

3.6 **Solapa *Structure* 2.** El bolsillo es mostrado como un conjunto de *alpha spheres* (polares-verdes- y apolares -blancas-), mientras que el ligando se encuentra representado por el tipo de átomo (carbonos, gris; nitrógenos, azul; oxígeno, rojo; azufre, amarillo y fósforo, violeta). En este caso el ligando corresponde a la droga isoniazida unida a NAD. 66

3.7 **Solapa *Metadata*.** La solapa *Metadata* permite visualizar la información de anotación funcional de UniProt como así también los perfiles de expresión génica en una diversidad de condiciones experimentales como son la exposición a ERON, hambruna, hipoxia y la infección de murinos. En el caso de la esencialidad si tiene un valor de ”1” o de ”YES” el gen/proteína es esencial. Para el caso de los experimentos de sobre expresión en diversas condiciones la interpretación resulta un poco más complicada, valores mayores a 2 se consideran que el gen en cuestión se encuentra sobre expresado mientras que valores menores a 0,5 se considera que el gen se encuentra reprimido en dicha condición. 67

3.8 ***Pipeline* de clasificación utilizado**. Para determinar la relevancia de una proteína (y su vía metabólica correspondiente) se utilizó la información depositada originalmente en TuberQ agregando los cómputos de vías metabólicas (proteínas que son cuellos de botella), expresión en condiciones que imitan la infección y de sensibilidad a ERON. Estos datos se encuentran actualmente depositados en TuberQ. 69

3.9 **Propiedades estructurales de *Inositol-3-Phosphate Synthase*** A) Vista del plegado de I3PS con el bolsillo drogable destacado en esferas rojas (PDBID 1GR0). B) Acercamiento del bolsillo drogable superpuesto con la estructura de NAD. Las cisteínas y tirosinas oxidables se encuentran dibujadas más gruesas. . 73

3.10 **Grafo de reacciones de la red metabólica de *Mycobacterium tuberculosis*.** Cada nodo representa una reacción predicha en el metabolismo de *Mtb*, y existe una arista entre nodos si el producto de una reacción es el sustrato de otra. El tamaño de los nodos representa el valor de *betweenness centrality* en el grafo de reacciones y puntuado primero de acuerdo con esta metrica. Los nodos en rojo representan cuellos de botella. Se destacan dos vías, *mycothiol biosynthesis* y *phosphatidyl-inositol biosynthesis* ambas teniendo altos puntajes según nuestro análisis. 75

3.11 **Via de síntesis del micotiol.** En rojo se muestran las proteínas más relevantes de la vía *ino1* y *mshB*. A un lado se muestra una representación del bolsillo drogable y de bolsillo se sabe drogable a través de la asociación por culpa al pertenecer a la misma familia de PFAM. FDN: falta de nutrientes 80

4.1 **Representación de guardas del plegado núcleo de las metiltransferasas dependientes de SAM.** Adaptado de (Martin, & McMillan, 2002) 94

4.2 95

4.3 **Mecanismo de reacción general propuesto para las CMAS de *Mycobacterium tuberculosis*.** La reacción se divide en dos pasos elementales: i) El ataque y formación del metil-carbocatión, común a todas las CMAS y ii) la resolución del mismo a distintos productos, propia de cada CMAS. 96

4.4 **Sistemas QM utilizado para cada una de las simulaciones:** A) Primer paso de la reacción en cmaA2. Se colocaron 4 *link atoms* para modelar la interfase entre los sistemas QM y MM, dos en los carbonos del SAM y dos en los carbonos de la olefina. B) Segundo paso de la reacción en cmaA2. Se colocaron 3 *link atoms*, dos en el carbocatión (en la misma posición que el paso anterior) y uno en el glutámico. C) Primer paso de la reacción en mmaA4. Se colocaron 2 *link atoms* en los carbonos del SAM. D) Segundo paso de la reacción en mmaA4. Se colocaron 2 li*nk atoms* en los glutámicos 126 y 129. Las lineas punteadas indican distancias utilizadas en las coordenadas de reacción, ver texto 98

4.5 **Mecanismo de reacción propuesto para cmaA2.** Las flechas indican el movimiento de los electrones. 99

4.6 **Mecanismo de reacción propuesto para mmaA4.** Las flechas indican el flujo de los electrones. 99

4.7 **Nomenclatura de la estructura secundaria de las *Cyclopropane Mycolic Acid Synthses***. Basado en (Huang et al., 2002) 101

4.8 **Alineamiento entre el molde y su secuencia objetivo para el modelado de: A)umaA, B) ufaA, C) mmaA1, D) mmaA3 y E) rv3720.** 102

4.9 **Estructura general de las CMAS (utilizando a cmaA2 como modelo - PDBID:1KPI) con una ampliación mostrando los componentes de su sitio activo.** 103

4.10 **Comparación estructural entre los modelos generados de las CMAS y los moldes utilizados:** A) pcaA-umaA, B) cmaA1-ufaA, C) cmaA1-mmaA1, D) cmaA1-mmaA3 y E) mmaA2 y rv3720. El código de color indica diferencias en el RMSD de la cadena principal siendo la escala de azul (bajo RMSD) a rojo (alto RMSD). 104

4.11 Alineamiento de secuencia de las proteínas con dominio CMAS (según PFAM) en el genoma de *M. tuberculosis*. Los colores corresponden a los utilizados por ClustalX. Los residuos involucrados en unión o que son parte del sitio activo se encuentran destacados en violeta. Las flechas indican posiciones importantes destacadas en el texto. 105

4.12 Arbol guía producido por UPGMA a partir del alineamiento utilizando como puntuación la matriz de sustitución BLOSUM62. 106

4.13 **Estructura de los sitios activos de las CMAS y del N-terminal de mmaA3 y mmaA4.** A) Estructura del sitio activo de cmaA2, mmaA1 y mmA4. En esferas y palitos se muestra la posición del ión bicarbonato en cmaA2. El resto de las posiciones corresponden a mmaA4, salvo el aspártico que pertenece a mmaA1. B) Estructura del N-terminal de mmaA4 (en gris) y mmaA3 (en azul). 107

4.14 **Estabilidad de cmaA2 durante 150ns de dinámica.** A) Calculo de RMSD de los átomos pesados de la cadena principal de cmaA2 respecto a la estructura cristalina de partida (PDBID: 1KPI). B y C) Vista de distintos ángulos de estructuras representativas a lo largo de la trayectoria de 150ns coloreadas desde azul (0ns) a blanco (75ns) a rojo (150ns). 110

4.15 **Perfil de energía libre de la reacción de metilación de una olefina en: A) En cmaA2 carbono C10, B) cmaA2 en carbono C11, C) cmaA2 en carbono C10 en mutante Y30A y D) en agua**. En azul se muestran los distintos trabajos y el naranja el perfil de energía libre calculado a partir de la inigualdad de Jarzynski. 111

4.16 **Resumen de los resultados energéticos del primer paso de la reacción de cmaA2.** 112

4.17 **Poblaciones de mulliken de la reacción de metilación de olefinas: A) cmaA2 en C10, B) cmaA2 en C11, C) cmaA2 Y30A en C10 y d) agua.** 113

4.18 **Distancias relevantes en la reacción de metilación de olefinas en:A) cmaA2 en C10, B) cmaA2 en C11, C) cmaA2 Y30A en C10 y d) agua.** 114

4.19 **Estructura del A) estado inicial, B) el estado de transición y C) el estado final del primer paso de formación del ciclopropano en cmaA2.** 115

4.20 **Perfil de energía libre de la reacción de resolución del carbocatión en: A) cmaA2 utilizando E142 como base adicional, B) sin E142 y C) agua.** En azul se muestran los distintos trabajos y el naranja el perfil de energía libre calculado a partir de la inigualdad de Jarzynski. 115

4.21 Resumen de los resultados energéticos del segundo paso de la reacción de cmaA2. 116

4.22 **Poblaciones de mulliken de la reacción de resolución del carbocatión en: A) cmaA2 utilizando E142 como base adicional, B) sin E142 y C) agua.** . . 116

4.23 **Distancias relevantes de la reacción de resolución del carbocatión en: A) cmaA2 utilizando E142 como base adicional, B) sin E142 y C) agua.** 117

4.24 **Estabilidad de mmaA4 durante 500ns de dinámica.** A y B) Vista de distintos ángulos de estructuras representativas a lo largo de la trayectoria de 500ns coloreadas desde azul (0ns) a blanco (250ns) a rojo (500ns). C) Calculo de RMSD de los átomos pesados de la cadena principal de mmaA4 respecto a la estructura cristalina de partida (PDBID: 3HA5) 117

4.25 **Perfiles de energía libre de la transferencia de metilo y formación del carbocatión al átomo A) C6 y B) C5 de la olefina modelo utilizada.** 118

4.26 **Distancias y cargas de la transferencia de metilo de SAM a una olefina en mmaA4 en sus dos carbonos.** A) Distancia transferencia a C5. B)Cargas transferencia a C5. C) Distancias transferencia a C6. D) Cargas transferencia a C6. 119

4.27 **Estructura del producto de reacción llevada a cabo por mmaA4.** 120

4.28 **Modelo propuesto del funcionamiento de las CMAS.** Arriba, las CMAS en general, abajo mmaA4. Al producirse la unión de los sustratos en mmaA4 no se produce un cierre del N-terminal mientras que en el resto de los miembros de la familia. La ampliación muestra que mmaA4 posee un sitio de aguas entre los dos glutámicos (un lugar donde es más probable encontrar aguas que en el seno del solvente) mientras que en cmaA2 este es mucho más débil. 123

5.1 **Fragmentos mimeticos del modo de unión de bicarbonato**. En la figura se encuentran señalados el perfil de interacciones de puente hidrógeno de cada fragmento y del anión bicarbonato. Don: Donor de puente hidrogeno, Acc:Aceptor de puente hidrógeno 131

5.2 **Sitio de unión de bicarbonato de umaA.** La numeración corresponde a umaA mientras que el bicarbonato proviene de una superposición de una estructura de cmaA2 (PDBId: 1KPH) con umaA. Las lineas punteadas corresponden a interacciones de puente hidrógeno. El anión bicarbonato se encuentra unido mediante cuatro puentes hidrógeno actuando en tres casos como aceptor (His(NE) 167, Cis 34 y Tir 231) y en un caso como donor (Glu 139). 134

5.3 **Estructuras de *docking* sobre umaA de los distintos fragmentos** A) Urea (ZINC36859431), B) Carbamato (ZINC02024095), C) Ácido metoxiacético (ZINC06059535) y D) Glicina (ZINC05374970). Las lineas punteadas indican interacciones de tipo puente hidrógeno. 136

5.4 **Estructura de las mejores drogas basadas en el motivo de unión de Urea.** Los identificadores son los siguientes: 1)ZINC03506439 2)ZINC36859431 3)ZINC12740738 4)ZINC09570734 5)ZINC06993185 6)ZINC02621554 7)ZINC11022390 8)ZINC06993187 9)ZINC08710095 10)ZINC15274986 137

5.5 **Estructura de las mejores drogas basadas en el motivo de unión de Carbamato.** Los identificadores son los siguientes: 1) ZINC02024095 2) ZINC03158901 3) ZINC04891055 4)ZINC04966565 5)ZINC02008606 138

5.6 **Estructura de las mejores drogas basadas en el motivo de unión de ácido metoxiacético.** Los identificadores son los siguientes: 1)ZINC06059535 2)ZINC17090657 3)ZINC16399519 4)ZINC17031713 5)ZINC04334497 6)ZINC12375164 7)ZINC04616853 8)ZINC00117863 9)ZINC17072207 10)ZINC00118984 139

5.7 **Estructura de las mejores drogas basadas en el motivo de unión de Glicina.** Los identificadores son los siguientes: 1)ZINC05374970 2)ZINC05269172 3)ZINC05374959 4)ZINC08382611 5)ZINC06659663 6)ZINC35655933 7)ZINC02049261 8)ZINC00063553 9)ZINC02357332 10)ZINC16267236 140

5.8 **Esquema de trabajo para el análisis de los mejores resultados de *docking* por dinámica molecular y MM-PBSA** 141

5.9 **Calculo de RMSD de las distintas drogas a lo largo de 20ns de dinámica molecular clásica** 142

5.10 **Interacciones de la foto promedio de la dinámica molecular clásica de: A) ZINC05374970 B)ZINC11022390 C) ZINC04891055 D) ZINC16399519** . . 147

5.11 **Comparación estructural del sitio activo de umaA (verde), cmaA1 (azul), cmaA2 (rojo) y pcaA (gris) superpuesto con el resultado de docking de ZINC06993185 sobre umaA** 149

6.1 **Sitio activo de umaA**. En la figura se muestra el sitio activo de umaA con todos los triptofanos de la proteína (en verde) y la mejor estructura de *docking* (realizado con rDock) de ANS (*8-Anilinonaphthalene-1-sulfonic acid*, amarillo) y análogos de ácidos micólicos (en violeta). 156

Tablas

3.1 **Proteínas de *Mtb* clasificadas de acuerdo a su *Druggability Score* (DS).** Los números entre paréntesis indican el número de proteínas que son esenciales como se definió en la sección métodos 69

3.2 **Número de proteínas sobre-expresadas en (1 a 4) condiciones tipo infección-** a) Números entre paréntesis corresponden solamente a proteínas drogables y esenciales b) El *Expression Score* (ES) describe el número de condiciones donde la proteína fue encontrada en sobre-expresión, desde 0 (la proteína no se sobre-expresa en condiciones tipo infección) a 4 (la proteína se sobre-expresa en las cuatro condiciones, hipoxia, hambruna, Estrés ERON e infección en ratones). . 71

3.3 **Proteínas de *Mtb* predichas como sensibles a ERON.** En a) proteínas altamente drogables b) Proteínas que están sobre expresadas si el puntaje de expresión es mayor a 3. Los números entre paréntesis corresponde únicamente a las proteínas drogables y esenciales. 72

3.4 Propiedades generales de la red metabólica de *Mtb*. 74

3.5 Blancos nuevos y revalidados encontrados en *Mtb* utilizando la drogabilidad estructural, el análisis de importancia metabólica y la información de expresión en condiciones de estrés de acuerdo a la priorización de TuberQ. 82

4.1 Resumen de las diferencias entre las distintas proteínas de la familia CMAS de *Mtb* 122

5.1 **Resultados de búsqueda de motivos estructurales en ZINC y de la búsqueda virtual en umaA** 135

5.2 **Resultados numéricos de MMPBSA de los compuestos completos.** Se presenta una decomposición en la componente polar de la interacción (Electroestatática y Solvatación de PB) y la componente no polar de la interacción (VdW y la contribución no polar de la energía libre de solvatación). Las unidades de todos los datos reportados en esta tabla son kcal/mol. 143

5.3 **Decomposición de la energía libre de unión estimada por MMPBSA para los átomos que pertencen a los cuatro fragmentos descriptos (Glicina, Urea, Metoxiacético y Carbamato).** En la quinta columna se presentan los resultados de ΔG de unión para cada fragmento relativizado al total de ese compuesto . . . 144

5.4 **Descomposición de la energía libre de unión estimada por MMPBSA para los grupos R**. Todas las energías se encuentran reportadas en kcal/mol. 145