



UNIVERSIDAD PERUANA  
**CAYETANO HEREDIA**

**PREDICCIÓN DE LA ACTIVIDAD  
ENZIMÁTICA DE PIRAZINAMIDASA  
WILD TYPE Y MUTADA DE  
*MYCOBACTERIUM TUBERCULOSIS*  
GUIADA POR MODELAMIENTO  
BIOMOLECULAR**

**TESIS PARA OPTAR EL GRADO DE  
MAESTRO EN INFORMÁTICA  
BIOMÉDICA EN SALUD GLOBAL CON  
MENCIÓN EN BIOINFORMÁTICA**

**GUSTAVO ENRIQUE OLIVOS RAMIREZ**

**LIMA – PERÚ**

**2021**



**ASESOR DE TESIS**

**PhD. Mirko Juan Zimic Peralta**

**JURADO DE TESIS**

DR. MIGUEL ANGEL CAMPOS SANCHEZ

PRESIDENTE

MG. JUAN CARLOS AGAPITO PANTA

VOCAL

MG. ANA CECILIA COLAROSI SALINAS

SECRETARIO (A)

## **DEDICATORIA**

A las personas que amo, Eva, Roberto y Fátima.

A los amigos que siempre me han apoyado, Kevin, Kewin, Geordano, y José.

A las personas que me brindaron su esencia, Fernando y José Luis.

A los estudiantes que sueñan con ser buenos profesionales.

## AGRADECIMIENTOS

Han sido muchas las personas quienes me han conducido hasta estas líneas. Dedicar un agradecimiento a cada una de ellas sería muy extenso, sin embargo, aquí vamos.

A mi asesor, el Dr. Mirko Zimic, por todas sus enseñanzas, apoyo, paciencia y las herramientas que ha puesto a disposición para la ejecución de esta investigación.

A los Dres. Cesar Carcamo y Patricia García, porque gracias a su trabajo es que la maestría ha podido realizarse.

A los excelentes docentes de la Facultad de Salud Pública, de manera especial a la Dra. Patricia Mallma, de quien pude aprender mucho.

A mis coordinadores del programa de maestría, Daniel y Paola, quienes desde el primer día nos han guiado y aconsejado.

A Georcki Geor Dano Ropón, por sus valiosas enseñanzas en el área de biofísica computacional.

A mis profesores de pregrado, quienes siempre me han aconsejado para bien, de manera especial a mi mentor Fernando Merino.

A mis compañeros de la maestría, especialmente a Kevin, Franklin, Mauro y Diane, por su valiosa amistad.

A mis compañeros del laboratorio, Stefany, Angelita, Gabriel, Mario, Manuel, Xiomara, Katherine, Akemi y María, de quienes pude aprender muchas cosas.

A las personas que a la distancia me han acompañado, Kewin, Claudina, Cleydi, Yaritza, Tania y Naty. Durante mis estudios nunca me dejaron solo.

A CONCYTEC por haberme financiado y permitido realizar la Maestría en Informática Biomédica en Salud Global.

A CONCYTEC-FONDECYT, por haberme confiado esta beca y financiar mis estudios de maestría. ¡Gracias por permitirme vivir esta gran aventura, Gracias!

Agradecer a CONCYTEC - FONDECYT por habernos proporcionado los fondos necesarios para la realización de esta tesis.

Al Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica CONCYTEC, a través del Fondo Nacional de Desarrollo Científico, Tecnológico y de Innovación Tecnológica FONDECYT.

## **FUENTES DE FINANCIAMIENTO**

**Este estudio fue realizado con un aporte de FONDECYT-CONCYTEC**

## TABLA DE CONTENIDOS

I. INTRODUCCIÓN .....	1
II. PLANTEAMIENTO DE LA INVESTIGACIÓN.....	4
2.1 MARCO TEÓRICO .....	4
2.1.1 Tuberculosis .....	4
2.1.2 Resistencia a Pirazidamida.....	5
2.1.3. Modelamiento Biomolecular.....	7
2.2 JUSTIFICACIÓN .....	11
2.3 OBJETIVOS .....	13
2.3.1 Objetivo General .....	13
2.3.2 Objetivos específicos .....	13
III. METODOLOGÍA .....	14
3.1 Diseño del estudio.....	14
3.2 Población y Muestra.....	14
3.3 Operalización de las variables.....	15
3.4 Recursos Computacionales y Software .....	16
3.6 Generación de mutaciones y refinamiento de las estructuras .....	16
3.7 Preparación del receptor.....	16
3.8 Preparación del ligando.....	17
3.9 Docking molecular .....	17
3.10 Simulación de dinámica molecular .....	18
3.11 Análisis de las dinámicas moleculares .....	19
3.12 Análisis de componentes principales .....	19
3.13 Parámetros geométricos y fisicoquímicos de las mutaciones .....	20
3.14 Análisis estadístico y modelo de predicción .....	20

3.15 Especificidad y sensibilidad.....	22
3.16 Renderizado y ploteo.....	23
IV. RESULTADOS.....	24
4.1 Parámetros cinéticos.....	24
4.2 Descripción del receptor y las estructuras mutadas .....	26
4.3 Descripción del ligando.....	28
4.4 Docking molecular .....	29
4.5 Dinámica molecular .....	35
4.6 Análisis de componentes principales (PCA – <i>essential dynamics</i> ) .....	42
4.7 Parámetros estructurales y fisicoquímicos .....	47
4.8 Predicción de la actividad enzimática .....	51
4.9 Especificidad y sensibilidad .....	54
V. DISCUSIÓN .....	56
VI. CONCLUSIONES .....	65
VII. RECOMENDACIONES .....	66
VIII. REFERENCIAS .....	67

## ÍNDICE DE FIGURAS

Figura 1. Incidencia de Tuberculosis en el mundo (por 100,000 habitantes) durante el 2017 (4). .....	4
Figura 2. Mecanismo molecular de la actividad de la pirazinamidasa a través de del medio intracelular e intercelular (22). .....	6
Figura 3. Workflow de los procedimientos bioinformáticos y de predicción. ....	22
Figura 4. A) Secuencia primaria de la enzima <i>PZasa</i> de <i>M. tuberculosis</i> . Las mutaciones reportadas están resaltadas en rojo. Algunos aminoácidos pueden sufrir más de una mutación (D12 y P62). B) Estructura tridimensional del cristal de la <i>PZasa</i> (58). Las hélices $\alpha$ están coloreadas en morado y las láminas $\beta$ en amarillo. ....	26
Figura 5. Distribución de las mutaciones puntuales caracterizadas en la enzima <i>PZasa</i> de <i>M. tuberculosis</i> . A) Mutaciones relacionadas con el centro de coordinación metálico. B) Mutaciones cercanas al sitio catalítico. C) Mutaciones en regiones distantes al sitio activo y de coordinación metálico. ....	27
Figura 6. Estructura molecular de la pirazinamida (izquierda) y nicotidamina (derecha). .....	28
Figura 7. Interacciones moleculares entre la pirazinamida y la enzima <i>PZasa</i> (superior). Modos de acoplamiento (inferior), en donde las moléculas de agua (W) están representadas como esferas. A) Acoplamiento del ligando hidratado en el sitio catalítico de la estructura <i>wild type</i> . B) Acoplamiento del ligando hidratado en el sitio catalítico de la estructura mutada C14G. ....	32
Figura 8. Interacciones moleculares entre la pirazinamida y la enzima <i>PZasa</i> (superior). Modos de acoplamiento (inferior), en donde las moléculas de agua (W) están representadas como esferas. A) Acoplamiento del ligando hidratado en el sitio catalítico de la estructura <i>wild type</i> . B) Acoplamiento del ligando hidratado en el sitio catalítico de la estructura mutada L4S. ....	33
Figura 9. Interacciones moleculares entre la pirazinamida y la enzima <i>PZasa</i> mutada en H57R (izquierda). Modos de acoplamiento de la droga y las moléculas de agua en el sitio activo (derecha). .....	35
Figura 10. Fluctuación del RMSD de la enzima <i>PZasa wild type</i> y con mutaciones cercanas al sitio de coordinación metálico. ....	36
Figura 11. Fluctuación del RMSD de la enzima <i>PZasa wild type</i> y con mutaciones cercanas al sitio catalítico. ....	37
Figura 12. Fluctuación del RMSD de la enzima <i>PZasa wild type</i> y con mutaciones lejanas al sitio activo y de coordinación metálico. ....	38

Figura 13. Fluctuaciones del RMSF de la enzima <i>PZasa wild type</i> y con mutaciones cercanas al sitio de coordinación metálico.....	39
Figura 14. Fluctuaciones del RMSF de la enzima <i>PZasa wild type</i> y con mutaciones en el sitio catalítico.....	40
Figura 15. Fluctuación del RMSF de la enzima <i>PZasa wild type</i> y con mutaciones lejanas al sitio activo y de coordinación metálico.....	41
Figura 16. Análisis de componentes principales de la enzima <i>PZasa wild type</i> y con mutaciones cercanas al sitio de coordinación metálico. El promedio de las mayores fluctuaciones está representado en líneas rojas, mientras que las fluctuaciones extremas están como líneas blancas.....	44
Figura 17. Análisis de componentes principales de la enzima <i>PZasa wild type</i> y con mutaciones en el sitio activo. El promedio de las mayores fluctuaciones está representado en líneas rojas, mientras que las fluctuaciones extremas están como líneas blancas. ....	45
Figura 18. Análisis de componentes principales de la enzima <i>PZasa wild type</i> y con mutaciones en el sitio activo. El promedio de las mayores fluctuaciones está representado en líneas rojas, mientras que las fluctuaciones extremas están como líneas blancas. ....	46
Figura 19. Fluctuación del volumen del sitio catalítico del modelo de la enzima <i>PZasa wild type</i> y de las mutaciones más representativas. ....	49

## ÍNDICE DE TABLAS

Tabla 1. Operalización de las variables del estudio. ....	15
Tabla 2. Parámetros cinéticos de la <i>PZasa</i> de <i>M. tuberculosis wild type</i> y mutada. ....	24
Tabla 3. Diferentes enfoques de docking molecular en el acoplamiento ligando-receptor. ....	29
Tabla 4. Parámetros estructurales de la <i>PZasa wild type</i> y mutada de <i>M. tuberculosis</i> . Los datos son obtenidos de las estructuras analizadas en el PCA. ...	47
Tabla 5. Parámetros fisicoquímicos de la <i>PZasa wild type</i> y mutada de <i>M. tuberculosis</i> . Los datos son obtenidos de las estructuras analizadas en el PCA. ...	49
Tabla 6. Análisis de regresión lineal múltiple entre los parámetros geométricos y fisicoquímicos para predecir la actividad enzimática de la <i>PZasa</i> de <i>M. tuberculosis</i> . ....	51
Tabla 7. Modelos para los parámetros cinéticos transformados y los parámetros de estabilidad, geométricos y fisicoquímicos. ....	53
Tabla 8. Predicción de la actividad enzimática con data no ingresada en el modelo. ....	54
Tabla 9. Matriz de confusión. ....	55

## RESUMEN

Pirazinamida es uno de los fármacos más usados en el tratamiento de la Tuberculosis. En la actualidad se reportan cepas resistentes a este compuesto, por tal motivo urge desarrollar métodos de detección de cepas sensibles y resistentes que ayuden a mejorar los tratamientos. Por ello, esta investigación tuvo como objetivo construir un modelo de predicción de la actividad enzimática de la enzima Pirazinamidasa *wild type* y mutada, usando data experimental de 35 mutaciones puntuales, en conjunto con métodos de simulación de acoplamiento molecular, dinámica molecular y un análisis de la dinámica esencial mediante PCA. Nuestros resultados han permitido identificar modificaciones a nivel de la estructura de la enzima *PZasa*, reportándose por primera vez un efecto de cierre y apertura del sitio activo como causa de las mutaciones. Asimismo, se ha logrado estimar las mayores fluctuaciones en estas conformaciones y asociarlas con parámetros geométricos y fisicoquímicos. Los mejores modelos de predicción se lograron por medio de una transformación logarítmica de los datos, para la constante catalítica ( $r^2=0.68$ ), actividad enzimática ( $r^2=0.67$ ), constante de Michaelis-Menten ( $r^2=0.65$ ) y eficiencia enzimática ( $r^2=0.29$ ). En la evaluación del modelo de la actividad enzimática, obtuvimos una sensibilidad de 55.56% y especificidad de 60.0%, lo cual sugiere que el modelo podría ser utilizado para predecir nuevas cepas resistentes, con cierto nivel de confiabilidad. Esta investigación representa el primer estudio que aborda la simulación y análisis de PCA para la enzima *PZasa*; además, los parámetros geométricos y enzimáticos pueden ser considerados en futuros trabajos que busquen mejorar la predicción de estos modelos.

**Palabras claves:** Tuberculosis, pirazinamida, modelamiento biomolecular, mutaciones.

## ABSTRACT

Pirazinamide is one of the most widely used drugs in the treatment of tuberculosis. Currently, resistant strains to this compound are reported, therefore, it is urgent to develop detection methods for sensitive and resistant strains that help to improve treatments. Therefore, this research aimed to build a prediction model of the enzymatic activity of the wild type and mutated pyrazinamidase enzyme, using experimental data of 35-point mutations, together with molecular docking simulation methods, molecular dynamics and an analysis of the essential dynamics by PCA. Our results have allowed us to identify modifications at the level of the *PZase* enzyme structure, reporting for the first time an effect of closure and opening of the active site as a cause of the mutations. We have also been able to estimate the major fluctuations in these conformations and associate them with geometrical and physicochemical parameters. The best prediction models were achieved by logarithmic transformation of the data for catalytic constant ( $r^2=0.68$ ), enzyme activity ( $r^2=0.67$ ), Michaelis-Menten constant ( $r^2=0.65$ ) and enzyme efficiency ( $r^2=0.29$ ). In the evaluation of the enzymatic activity model, we obtained a sensitivity of 55.56% and specificity of 60.0%, suggesting that the model could be used to predict new resistant strains, with a certain level of reliability. This research represents the first study that addresses the simulation and analysis of PCA for the *PZase* enzyme; furthermore, the geometric and enzymatic parameters can be considered in future work that seeks to improve the prediction of these models.

**Keywords:** Tuberculosis, pyrazinamide, biomolecular modeling, mutations.

## I. INTRODUCCIÓN

A lo largo de la historia de la humanidad, los microorganismos patógenos han sido responsables de diversas enfermedades, causando la muerte de millones de seres humanos (1). Resulta difícil calcular los impactos sociales y económicos que han ocasionado y aunque en la actualidad los avances en medicina y tecnología médica permiten combatir a muchos de ellos, algunos agentes infecciosos continúan representando un serio peligro para la salud global (2).

La Tuberculosis (TB), causada por *Mycobacterium tuberculosis*, es una de las enfermedades infecciosas más importantes en todo el mundo, debido a su alta prevalencia, morbilidad y mortalidad (3). En el año 2018, se reportaron más de 10 millones de casos de infecciones confirmadas y 1.6 millones de muertes por esta enfermedad; además, se ha determinado que su prevalencia está relacionada con el nivel socio-económico, ocurriendo con más frecuencia en países pobres y en desarrollo (4). Por tal motivo, diversas iniciativas se desarrollan en todo el mundo con la finalidad de erradicarla (5). A pesar de ello, nuevos problemas relacionados a su tratamiento emergen, lo que conlleva a una necesidad imperativa de abordarlos rápidamente.

En el Perú, se reporta un incremento en los casos de infecciones de TB multi-drogo resistente, los cuales representan cerca del 79% del total de casos de TB-MDR y el 70% de los casos de TB extensamente resistente (TB-XDR) en la Región de las Américas (6). Entre los años 2013 y 2015, el 57.7% de los casos presentaron resistencia a 5 y 7 drogas y 41.6% presentaron resistencia a 8 y 10 drogas de primera y segunda línea (7). Esta situación conlleva a que los tratamientos sean más

prolongados y además que los antimicrobianos tradicionalmente usados presenten cada vez menos efectividad (8).

La droga Pirazinamida (PZ) es uno de los fármacos más importantes en el tratamiento contra la TB y representa el único tratamiento capaz de eliminar a *M. tuberculosis* en estado de latencia (9). El uso de esta droga ha demostrado resultados positivos para reducir el tiempo de tratamiento de TB, de 9 a 6 meses (10), sin embargo, desde hace algunos años, ciertas cepas de *M. tuberculosis* han desarrollado resistencia a este compuesto (11).

Recientemente se ha descrito que el sobre uso de antibióticos, o el mal uso de estos, ejerce una mayor presión evolutiva sobre nuevas cepas (12). Esta situación promueve su prevalencia, esparcimiento y en un futuro cercano el surgimiento de nuevos patógenos altamente resistentes como un fenómeno global (13). Cabe mencionar que, la resistencia antimicrobiana se ha convertido en uno de los mayores problemas en salud pública a nivel mundial. A pesar del desarrollo de antibióticos en las últimas décadas, se reporta un incremento de casos de infecciones de bacterias multi-drogo resistentes y extremadamente resistentes (14).

Se ha postulado que la función enzimática de las proteínas está determinada por la conformación que adquiere durante su plegamiento (15). Es así que, los cambios conformacionales de las enzimas, como resultado de una dinámica diferente en aquellas estructuras mutadas, modifican su actividad y pueden conllevar a la resistencia de algunos fármacos (16). En ese sentido, determinar la relación estructura/función de las moléculas continúa siendo un gran reto para la medicina.

En la actualidad, el desarrollo de métodos teóricos para el modelamiento de sistemas biológicos ha permitido entender la relación de la dinámica de proteínas y también determinar su posible actividad mediante métodos de simulación bajo condiciones fisiológicas (17). En ese sentido, predecir el comportamiento de las moléculas mediante métodos de simulación permite entender la actividad de los fármacos sobre las moléculas de targets (18). A pesar de ello, poco se ha explorado en el desarrollo de modelos predictivos que permitan determinar la actividad enzimática basado en datos de simulaciones computacionales, lo cual resulta interesante ya que apertura las posibilidades de establecer mejoras en los tratamientos basado en un criterio de selección de fármaco y actividad catalítica.

La bioinformática viene surgiendo como una herramienta para abordar diferentes problemas en el área biomédica (19). Esta disciplina se basa en el uso de base de datos, servidores, algoritmos, entre diversos recursos computacionales, para resolver preguntas de investigación desde un enfoque biológico (20). El desarrollo de esta herramienta ha permitido simular procesos biológicos para predecir la conducta de fármacos, proporcionando información relevante para entender sus mecanismos de acción sobre las moléculas y mejorar los tratamientos (21).

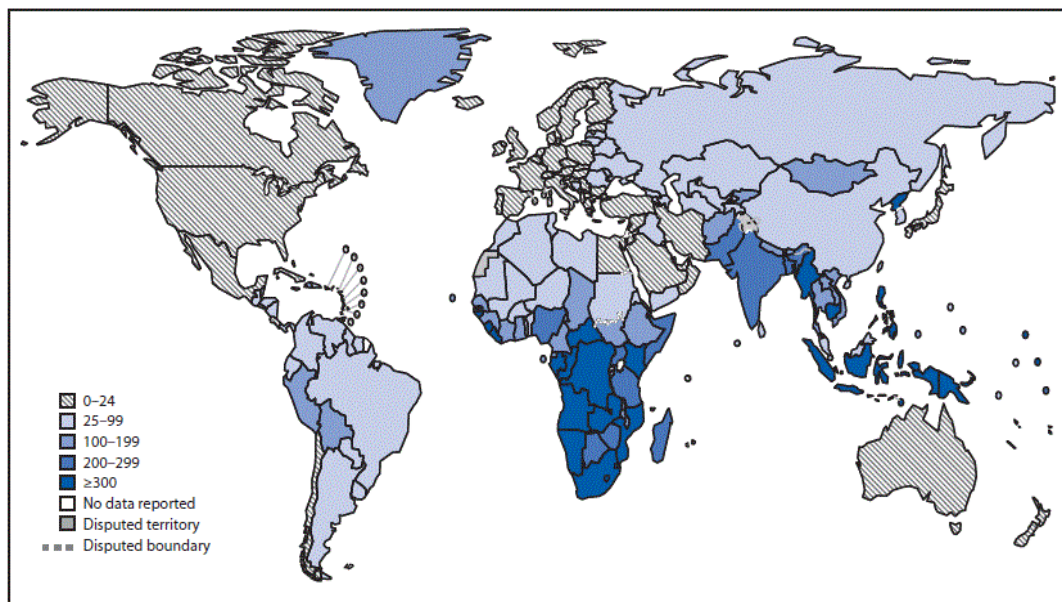
Bajo este contexto, la presente investigación tuvo como objetivo proponer un modelo de predicción de la actividad de enzimática de *PZasa*, basado en datos obtenidos mediante análisis de dinámica molecular. Para ello, se empleó un *data set* de mutaciones reportadas experimentalmente en conjunto con métodos bioinformáticos y de simulación computacional.

## II. PLANTEAMIENTO DE LA INVESTIGACIÓN

### 2.1 MARCO TEÓRICO

#### 2.1.1 Tuberculosis

La tuberculosis es una de las enfermedades más antiguas de la humanidad. Ha generado millones de muertes a lo largo de su historia y en la actualidad continúa siendo una de las principales causas de muerte en el mundo. Es una enfermedad infecciosa causada por *M. tuberculosis*, un bacilo extremadamente difícil de combatir. Se encuentra distribuida a nivel mundial y afecta mayormente a las poblaciones de menor recurso económico.



**Figura 1.** Incidencia de Tuberculosis en el mundo (por 100,000 habitantes) durante el 2017 (4).

El bacilo que ocasiona esta enfermedad invade el tracto respiratorio en los humanos, principalmente los pulmones, ocasionando inflamaciones en los alvéolos, lo cual produce acumulación de fluidos y purulencias. Su contagio se da por vía aerógena,

mediante pequeñas gotas de agua que son expulsadas por pacientes infectados con el bacilo; otras formas de infección pueden darse por vías urogenitales, transplacentaria y por inoculación, sin embargo, estas son menos frecuentes. A pesar de que una persona llegue a adquirir la enfermedad, solo 10% de la población infectada llegará a desarrollar la enfermedad, dado que en la mayoría de los casos el bacilo permanecerá en estado de latencia.

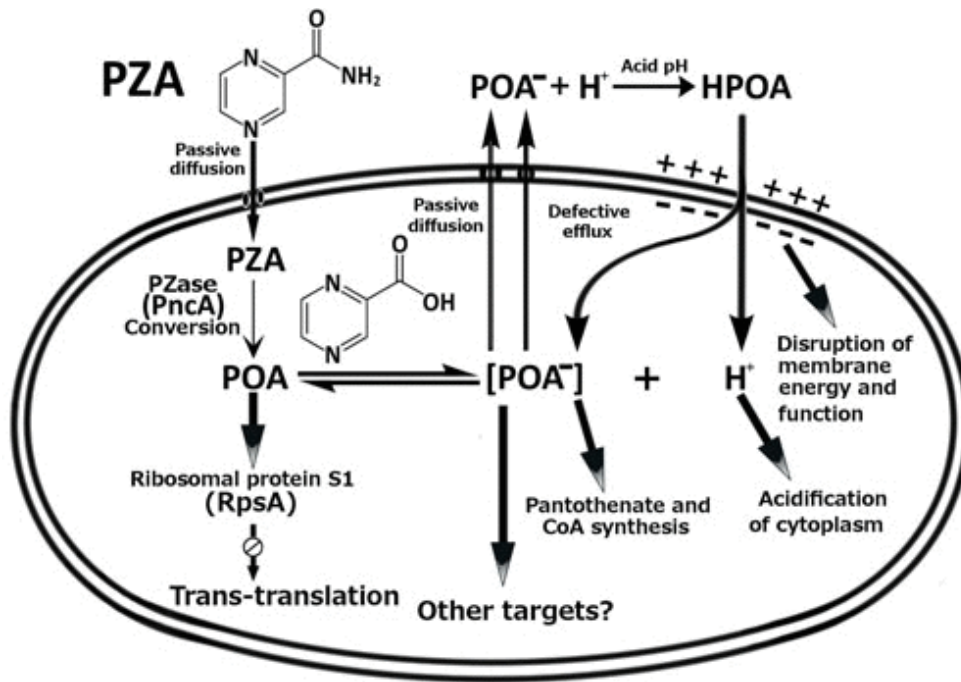
### **2.1.2 Resistencia a Pirazidamida**

Pirazinamida es uno de los principales fármacos empleado en el tratamiento de la TB, el cual ha contribuido notablemente a combatir las infecciones por *M. tuberculosis*, aún en estado de latencia, y reducir el tiempo de su tratamiento (de 12 a 6 meses). A pesar de ello, reportes a nivel mundial indican que diversas cepas de *M. tuberculosis* están desarrollando resistencia a este fármaco.

La Pirazinamidasa es la enzima que hidroliza a Pirazinamida, transformándola en su forma activa, ácido pirazinóico (POA). Esta enzima produce POA a un pH ácido y actúa únicamente en presencia de un cofactor metálico Zinc ( $Zn^{+}$ ) o Hierro ( $Fe^{+}$ ). Mutaciones puntuales en esta enzima, como L19R, R140H y E144K, han demostrado generar pérdida en la energía de unión con la pro-droga, lo cual explicaría la resistencia a este fármaco.

La representación del mecanismo de acción de Pirazinamida en *M. tuberculosis* se explica en la Fig. 2. La pro-droga ingresa mediante difusión pasiva (VPZA) al citoplasma de la bacteria, en donde es transformada en POA<sup>-</sup> por la enzima PZasa. Posteriormente, el POA<sup>-</sup> es liberado al medio extracelular por una bomba de flujo (VPOA), en donde es protonado a POA<sup>+</sup> y reingresado a la bacteria por una

diferencia del gradiente de concentración, interfiriendo así con otros procesos metabólicos de la bacteria.



**Figura 2.** Mecanismo molecular de la actividad de la pirazinamidasa a través de del medio intracelular e intercelular (22).

La resistencia a Pirazinamida ha sido relacionada con el gen *PncA*, el cual codifica a la enzima Pirazinamidasa (*PZasa*), principalmente con mutaciones que conducen cambios de aminoácidos en el sitio activo y de coordinación de la proteína *PZasa*. Estudios genómicos han corroborado que la resistencia ocurre únicamente en secuencias mutadas, no encontrándose mutaciones en las cepas sensibles (23). Las mutaciones ocasionan pérdida de afinidad por la droga y con el cofactor metálico, siendo estos eventos los más importantes para la resistencia (24). Asimismo, la pérdida de la estabilidad en la dinámica del complejo (fármaco-receptor) ha sido atribuido a las mutaciones (25).

Es importante señalar que Pirazinamida posee otros posibles efectos contra *M. tuberculosis*, entre los cuales: interfiere con el potencial de membrana y la producción de energía (22), inhibe *RpsA* (proteína ribosoma S1) relacionada con la trans-traslación (26), y *PanD* (aspartato descarboxilasa), involucrada en la síntesis de pantoténato y co-enzima A (27). Es por ello que el uso de esta droga resulta imprescindible en el tratamiento de la TB, asimismo es vital entender los mecanismos moleculares que ocurren en cepas *wild type* y mutantes.

### **2.1.3. Modelamiento Biomolecular**

#### **A) Docking molecular**

El docking molecular (acoplamiento molecular) es un método para determinar el modo de acoplamiento de un fármaco sobre una proteína target y mediante el cual es posible inferir la posible actividad del mismo (28). Este método permite la evaluación de la conformación y orientación de uno o varios ligandos (drogas) en el sitio activo de una proteína. Este método se aplica mediante el uso de dos tipos de algoritmos, uno de pasos múltiples que calcula la suma de energías en un complejo (proteína-ligando), el cual asigna un valor de puntuación para cada conformación con las que se obtienen las mejores conformaciones, y un segundo algoritmo denominado algoritmo Lamarckiano (*genetic algorithm*) el cual determina las mejores conformaciones (poses) de los ligandos (29).

El algoritmo genético se basa en la teoría de la selección natural para la búsqueda de las mejores conformaciones del ligando en el sitio activo de la proteína, generando poblaciones de conformómeros de los ligandos (aleatoriamente) y seleccionando aquellos con mejor energía de interacción (30). Mientras que la

función de *scoring* es una descripción matemática (ecuación 1) para el cálculo de la energía de unión basado en los términos descritos en la ecuación 2 (31).

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{unbound}^{P-P}) \quad (1)$$

$$+ (V_{bound}^{P-L} - V_{unbound}^{P-L} + \Delta_{sconf})$$

$$V = W_{vdw} \sum_{i,j} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + W_{hbound} \sum_{i,j} E(t) \left( \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + \quad (2)$$

$$W_{elec} \sum_{i,j} \left( \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} \right) + W_{sol} \sum_{i,j} (S_i V_j + S_j V_i) e^{(-r_{ij}^2/2\sigma^2)}$$

## B) Dinámica molecular

Las proteínas son moléculas relevantes en el estudio de fármacos, dado que ellas participan en procesos metabólicos vitales que pueden ser inducidos o inhibidos para tratar diversas enfermedades (32). Su función se encuentra íntimamente relacionada con su estructura y la dinámica del movimiento de sus átomos, asimismo, la actividad biológica de estas moléculas depende de las interacciones que ocurren entre ellas (33).

En la actualidad, se han desarrollado modelos teóricos para la predicción del comportamiento de las proteínas, mostrando gran precisión y confiabilidad de los resultados debido a que su descripción se basa en la resolución del movimiento de sus átomos mediante principios termodinámicos (34). Estos modelos están implementados sobre algoritmos computacionales y son empleados en diversas disciplinas (35).

Uno de los métodos más utilizados para simular el comportamiento de las moléculas es la simulación de dinámica molecular (SMD). Esta técnica es aplicada

para la investigación en campos como la bioquímica estructural, enzimología, biofísica, biología molecular, biotecnología y química farmacéutica (36). Mediante simulación de dinámica molecular es posible observar los eventos que ocurren durante la formación de un complejo molecular (proteína-ligando), en el espacio-tiempo, y que usualmente no pueden ser determinados experimentalmente (37).

El comportamiento de una proteína en el tiempo puede ser descrita bajo un enfoque mecánico clásico, el cual aproxima el movimiento de la molécula y sus átomos. En una simulación de dinámica molecular, la descripción de del movimiento en el sistema es resuelta mediante la ecuación (3) del movimiento de Newton (38):

$$m_i \frac{\partial r_i}{\partial t_i} = F_1; i = 1 \dots n \quad (3)$$

En donde, “m” y “x” representan la masa y la posición de la partícula, respectivamente, y “t” es el tiempo. El término  $F(x)$  representa la fuerza de restauración mecánica efectiva (fuerza media) que actúa sobre la partícula,  $f \, dx/dt$  es la fuerza de fricción relacionado al movimiento de la partícula en el medio y  $R(t)$  las fuerzas de fluctuación aleatorias. Mediante esta ecuación, las posiciones y las velocidades, como una función de tiempo, son calculadas para todos los átomos que conforman la molécula.

Los métodos de simulación computacional han contribuido a dilucidar procesos biológicos a un nivel atomístico, permitiendo la descripción de la termodinámica de las proteínas y sus propiedades en el tiempo. El entendimiento de estos comportamientos permite la exploración de la actividad de diversos compuestos, entre ellos los fármacos como Pirazinamida, y así anticipar la efectividad de su uso.

Con la finalidad de realizar estudios de simulación de dinámica molecular, se dispone del cristal de la enzima Pirazinamidasasa de *M. tuberculosis* (PDB ID: 3PL1), con una resolución de 2.2 Å, el cual ha provisto importantes detalles para el entendimiento de la acción de Pirazinamida. Asimismo, su cristalización posibilita realizar estudios de mutaciones y su impacto en la resistencia antimicrobiana. Estudios preliminares de simulación de dinámica molecular en *PZasa* confirman modificaciones estructurales ocasionadas por mutaciones en el gen *PncA*, las cuales reducen el volumen del *binding pocket* (sitio de unión) y modifican la estabilidad en la formación del complejo, lo cual conduce a la pérdida de la afinidad por la droga.

### **C) Dinámica molecular esencial o análisis de componentes principales**

El análisis de la dinámica molecular esencial (*principal components analysis*, PCA) es un método para determinar las fluctuaciones de mayor amplitud que representan la mayor contribución en la función de una proteína (enzima). Este análisis determina los movimientos significativos de una molécula, en función del tiempo, mediante un análisis de los datos obtenidos de una simulación de dinámica molecular. El método se basa en un análisis de covarianza a partir de una matriz generada de la diagonalización de las trayectorias, considerando únicamente los carbonos alpha (C $\alpha$ ) y, posteriormente, los *eigenvalues* y *eigenvectors* son calculados tomando las trayectorias. Se considera que mediante este procedimiento es posible relacionar la dinámica (movimiento) de las moléculas con su función enzimática.

## **D) Predicción de la actividad Enzimática**

El entendimiento y la predicción de la actividad enzimática de las proteínas como resultado de sus modificaciones estructurales y cinéticas permanece como un reto para la ciencia. A pesar de ello, algunos modelos para la predicción enzimática mediante colecciones de datos y algoritmos de entrenamiento han demostrado resultados alentadores. Asimismo, algunas investigaciones señalan que es posible correlacionar parámetros estructurales de las proteínas con la función enzimática (39), sin embargo, estos modelos necesitan ser validados con datos experimentales.

Diferentes modelos de predicción, como los que usan regresión lineal, son utilizados frecuentemente en diversas áreas de estudios. La disponibilidad de datos experimentales de la actividad enzimática de PZasa de *M. tuberculosis*, permite correlacionar estos datos con los obtenidos de la simulación biomolecular. En ese sentido, resulta posible proponer un modelo teórico para la predicción de la actividad enzimática de este target.

## **2.2 JUSTIFICACIÓN**

En la actualidad, la resistencia antimicrobiana es una de las mayores preocupaciones para todos los países del mundo, debido a que se reporta un aumento en el número de cepas resistentes a diversos fármacos, situación que limita los tratamientos. Ello representa un reto para la búsqueda de alternativas con miras a disminuir la resistencia antimicrobiana, previniendo una situación de pandemia global por bacterias multi-drogo-resistentes.

Si bien se ha logrado determinar mutaciones (a nivel genómico) que causan resistencia, resulta necesario conocer y describir los mecanismos moleculares y

termodinámicos asociados a las mutaciones a nivel estructural, los cuales puedan servir para determinar la modificación de la actividad enzimática de los *targets* (blancos farmacológicos). Este conocimiento permitirá proponer modelos para la predicción enzimática, los cuales contribuyan en el diseño de mejores tratamientos mediante una selección oportuna de fármacos para cada cepa sensible y por ende disminuir la presión de selección que ocasiona la drogo-resistencia.

Los cambios conformacionales en las estructuras proteicas representan un concepto clave para comprender la dinámica y actividad (función) de estas moléculas. Estos cambios pueden ser estudiados y entendidos mediante simulación biomolecular, dado que proveen de una vista a nivel atómico de los eventos que ocurren en las moléculas. Ello representa una oportunidad para la aplicación de estas metodologías en el estudio de las mutaciones y la drogo-resistencia.

Asimismo, los métodos de simulación computacional permiten describir sistemas biológicos con mayor precisión, contribuyendo al conocimiento para la mejora de fármacos para el tratamiento de la TB. En ese sentido, determinar los efectos de las mutaciones sobre la actividad enzimática de los *targets* con la finalidad de aplicar tratamientos específicos para cada cepa contribuirá a combatir la resistencia de *M. tuberculosis*.

## 2.3 OBJETIVOS

### 2.3.1 Objetivo General

Determinar la actividad enzimática de *PZasa wild type* y mutada de *M. tuberculosis* mediante un esquema de modelamiento molecular, que incorpora simulaciones de dinámica molecular y un análisis de componentes principales (PCA).

### 2.3.2 Objetivos específicos

- Determinar variaciones en la energía de interacción ( $\Delta\Delta G$ ) entre *PZasa wild type* y mutantes de *M. tuberculosis* con pirazinamida mediante docking molecular.
- Determinar la dinámica de la interacción entre *PZasa wild type* y mutantes de *M. tuberculosis* mediante una simulación de dinámica molecular y análisis de PCA.
- Determinar la sensibilidad y especificidad para predecir la actividad enzimática de *PZasa* mediante los datos obtenidos del modelamiento biomolecular.

## 2.4 Hipótesis de trabajo

La actividad enzimática de la enzima *PZasa wild type* y mutada de *M. tuberculosis* puede ser predicha mediante modelamiento biomolecular.

### III. METODOLOGÍA

#### 3.1 Diseño del estudio.

La investigación fue un estudio descriptivo-analítico. Este estudio aborda principalmente simulaciones de acoplamiento molecular, dinámica molecular, análisis de componentes principales, cálculos de parámetros geométricos y fisicoquímicos; además de construcción de modelos mediante regresión lineal múltiple y transformaciones de datos.

#### 3.2 Población y Muestra

Se utilizó un conjunto de datos de mutaciones en la enzima *PZasa*, asociadas con la resistencia a la Pirazinamida. La base de datos de las mutaciones ha sido construida y validada experimentalmente mediante ensayos de actividad cinética produciendo *PZasa* recombinante, purificándola y determinando sus respectivos parámetros cinéticos (24), en el Laboratorio de Bioinformática y Biología Molecular de los Laboratorios de Investigación y Desarrollo de la UPCH. Dichas mutaciones cuentan con parámetros enzimáticos como  $K_{cat}$ ,  $K_m$ ,  $Eff$  y actividad enzimática ( $Act.$ ), para 35 mutaciones puntuales (resistentes) y dos *wild type* (sensibles). Cabe señalar que los datos debido a la forma en la que se obtuvieron los datos experimentales, estos están divididos en dos grupos (*batches*), el primero cuenta con 29 datos (28 mutantes y 1 *wild type*) y el segundo de 8 (7 mutantes y 1 *wild type*). Estos parámetros han sido utilizados para los análisis de regresión con los datos obtenidos mediante simulación biomolecular.

### 3.3 Operalización de las variables

Las variables del estudio y sus respectivos indicadores se describen en la tabla 1.

Estas variables se componen por los datos obtenidos en las simulaciones (bioinformáticas) y los parámetros cinéticos (experimentales).

**Tabla 1.** Operalización de las variables del estudio.

	<b>Variables</b>	<b>Definición</b>	<b>Indicadores</b>
<b>Bioinformáticas</b>	Energía de interacción	Cantidad de energía requerida para separar dos moléculas en un sistema.	Energía de interacción (Kcal mol <sup>-1</sup> )
	Potencial electrostático	Distribución de las cargas eléctricas en la superficie de las proteínas.	Carga de superficie, Electronegatividad (pI)
	Volumen del sitio activo	Volumen que conforma la cavidad del sitio de unión.	Volumen Angstrom (Å <sup>3</sup> )
	Distancias	Distancia entre residuos importantes para la actividad enzimática.	Distancia (Å)
	SASA	Superficie accesible al solvente	Área (Å <sup>2</sup> )
	RMSD	Desviación cuadrática media de los carbonos alfa de la proteína.	RMSD (Å)
	RMSF	Desviación cuadrática media de cada residuo de la proteína.	RMSF (Å)
	Radio de giración	Describe la forma en la cual el área transversal de la proteína se distribuye alrededor de su eje centroidal.	Rg (Å)
	Índice de Inestabilidad	El índice de inestabilidad proporciona una estimación de la estabilidad de una proteína.	II
<b>Experimentales</b>	Constante Catalítica o número de recambio	Número de moléculas de sustrato convertida por la enzima en producto por unidad de tiempo.	<i>Kcat</i>
	Constante de Michaelis-Menten	Concentración del sustrato en el cual la ratio inicial alcanza la mitad de la velocidad máxima.	<i>Km</i>
	Eficiencia enzimática	Mediad de la eficiencia de una enzima.	<i>Eff</i> ( <i>Kcat</i> / <i>Km</i> )
	Actividad Enzimática	Cantidad de enzima que cataliza la conversión de 1 μmol de sustrato en un minuto.	<i>Act.</i> (μmol min <sup>-1</sup> )

### 3.4 Recursos Computacionales y Software

Se trabajó en un servidor con las siguientes características: procesador Intel core i7 de 8 núcleos, 32 GB RAM, HD 1TB, RTX TITAN 11 GB DDR5 (3584 Cuda Cores) y RTX 2080 12 GB GDDR6 (4350 Cuda Cores), bajo un entorno de Ubuntu v 18.04.03 LTS.

### 3.5 Bases de datos

El cristal de la proteína Pirazinamidasa (GenBank: BAO48146.1) (40) se obtuvo de la base de datos Protein Data Bank (PDB ID: 3PL1, Resolución: 2.2 Å) (41), en formato PDB. Esta estructura se utilizó debido a que es la única proteica resuelta y disponible para análisis *in silico*. Todas las mutaciones se generaron a partir de este cristal. Por otro lado, la estructura química de la droga Pirazinamida se obtuvo del repositorio PubChem (PubChem CID: 1046) en formato \*.SDF.

### 3.6 Generación de mutaciones y refinamiento de las estructuras

De las mutaciones disponibles en la base de datos del laboratorio, 36 mutaciones en la enzima *PZasa* fueron consideradas. Para mutar y obtener cada mutante, se utilizó un *script* en Python (Anexo 1) que genera una mutación puntual con el software Modeller v9.21 (42), luego la refina mediante un algoritmo de gradiente conjugada. Las estructuras mutadas fueron guardadas en formato \*.PDB para su posterior uso.

### 3.7 Preparación del receptor.

Las moléculas de agua que fueron co-cristalografiadas con la enzima fueron eliminadas tanto para el *wild type* y las mutantes, mediante el software Autodock Tools v. 1.5.7 (30). Posteriormente, hidrógenos polares y cargas *Gasteiger* fueron

añadidos conforme a lo recomendado. Las estructuras fueron transformadas a formato \*.PDBQT para los análisis de docking molecular.

### 3.8 Preparación del ligando

La pirazinamida, obtenida en formato \*.SDF, se convirtió a formato \*.PDB generando coordenadas 3D, a pH=7.4, adicionando hidrógenos polares, todo ello mediante un *bash script* (Anexo 2) que usa el software OpenBabel v2.4.1 (43). Asimismo, en este procedimiento se ejecutó un paso de optimización geométrica mediante minimización de energía con MMFF94. La estructura final fue transformada a \*.PDBQT con un *script* provisto por los paquetes de Autodock Tools.

### 3.9 Docking molecular

Los análisis de docking molecular se realizaron con la finalidad de determinar el efecto de las mutaciones sobre la interacción fármaco-receptor (PZasapirazinamida). Es por ello que, cada mutante y *wild type* fue simulada mediante docking molecular con el software Autodock v4 (44). Mediante este software se emplearon 5 metodologías de docking para modelar el acoplamiento de la droga con las estructuras mutadas: 1) docking sin ion metálico, 2) docking hidratado sin ion metálico, 3) docking con ion metálico, 4) docking hidratado con ion metálico y 5) docking con ion cargado (+2). El box de simulación fue una caja de 60<sub>x</sub>, 60<sub>y</sub>, 60<sub>z</sub>; la cual incluye a los aminoácidos Asp8, Lys96 y Cys138, los cuales han sido descritos como los más importantes para la catálisis enzimática (41). El software utiliza un algoritmo genético o algoritmo Lamarckiano para la determinación de las mejores conformaciones (poses) entre los complejos. Los parámetros de docking fueron: tamaño de población de 350, número de evaluaciones 2500000, número de

generaciones de 27000, ratio de mutación 0.02, cross over ratio 0.8, run=100. Las energías se seleccionaron tomando aquellas que tenían mayor número de individuos para un *cluster*, y fueron estimadas en Kcal mol<sup>-1</sup>.

### **3.10 Simulación de dinámica molecular**

Las simulaciones se realizaron usando el software GROMACS v 2018.3 (45,46), en su configuración GPU. Cada una de las estructuras mutadas y el *wild type* fueron sometidas a 500 ns de simulación, tomando únicamente a la proteína, con la finalidad de evaluar las modificaciones estructurales producto de las mutaciones. Se usó el campo de fuerza OPLSAA/FF.

El procedimiento para realizar la simulación de dinámica molecular consistió de los siguientes pasos: 1) se generó la topología del sistema con el módulo GMX2PDB de GROMACS, 2) el sistema es solvatado con moléculas de agua del modelo SPCE, en una caja cubica de 1.0 nm de distancia entre la proteína y la pared de la caja, 3) el sistema es neutralizado agregando iones del tipo cloro (Cl) y sodio (Na), 4) el sistema es minimizado energéticamente usando el algoritmo de pasos descendientes con 5000 pasos con la finalidad de eliminar malos contactos en el sistema, 5) se realiza un paso de equilibración de la temperatura utilizando un ensamble isocórico-isotérmico a una temperatura de 310.15 K (37°C) y el termostato de Berendsen, seguido de un paso de equilibración de la presión con un ensamble isobárico-isotérmico a una presión de 1 bar, usando el baróstato de Parrinello-Rahman, 6) finalmente, la dinámica es producida con pasos de integración de 2 fs, bajo condiciones de temperatura y presión constante. Durante la producción, el algoritmo de LINCS fue usado para restringir todos los enlaces, mientras que el

algoritmo de *Edwad's Mesh particle* para aproximar las interacciones de largo rango.

### **3.11 Análisis de las dinámicas moleculares**

Las trayectorias fueron centradas y se eliminaron los movimientos asociados al desplazamiento rotacional y traslacional de cada sistema, con la finalidad de obtener aquellos movimientos asociados a las vibraciones propias de la enzima. Con estas trayectorias, se construyeron los plots de RMSD (*root mean square deviation*) y RMSF (*root mean square fluctuation*), con los módulos *gmx\_rms* y *gmx\_rmsf* de GROMACS.

### **3.12 Análisis de componentes principales**

La función de una proteína a menudo está íntimamente ligada a su dinámica conformacional, sin embargo, no siempre es sencillo extraer los movimientos funcionalmente relevantes. Una de las principales razones de esto es que en una dinámica molecular observamos que todo se mueve al mismo tiempo. Tanto las fluctuaciones locales como los movimientos colectivos ocurren simultáneamente, lo que dificulta distinguir los dos tipos de movimiento entre sí. Un análisis de componentes principales (PCA) puede ayudar en tales casos, ya que permite filtrar los movimientos globales, colectivos (a menudo lentos) de los movimientos locales y rápidos. Por tal motivo, en nuestro estudio empleamos este análisis para identificar los mayores cambios estructurales de las estructuras *wild type* y mutadas.

El análisis de componentes principales (PCA) se realizó usando el módulo *gmx\_covar* y *gmx\_anaeig* de GROMACS. Las trayectorias de los PCA se

almacenaron en formato \*.PDB para su posterior análisis. Este análisis se realizó considerando únicamente los últimos 200 ns de las dinámicas moleculares, debido a que presentan las menores fluctuaciones en las simulaciones.

### **3.13 Parámetros geométricos y fisicoquímicos de las mutaciones**

Para este análisis, se tomó el promedio del PCA con las mayores fluctuaciones. El volumen del sitio activo se determinó utilizando el plugin PyVol (47) implementado en PyMol v. 2.4.1. Los volúmenes se determinaron en Å<sup>3</sup>. La superficie accesible al solvente se determinó con un *TCL script* (Anexo 3) que utiliza VMD v1.9.4a43. Además, se determinaron las distancias entre los residuos importantes (His57 y Asp49) para el centro de coordinación metálico, el RMSD de cada una de las estructuras, los RMSF de los residuos 8, 49, 57, 71, 133, 137 y 138, los cuales están relacionados a la coordinación del metal y la actividad catalítica, el radio de giración (Rgx), el peso molecular, punto isoeléctrico (PI), índice de inestabilidad (II), e índice alifático (AI), estimado en el servidor de ProtParam (<https://web.expasy.org/protparam/>) (48).

### **3.14 Análisis estadístico y modelo de predicción**

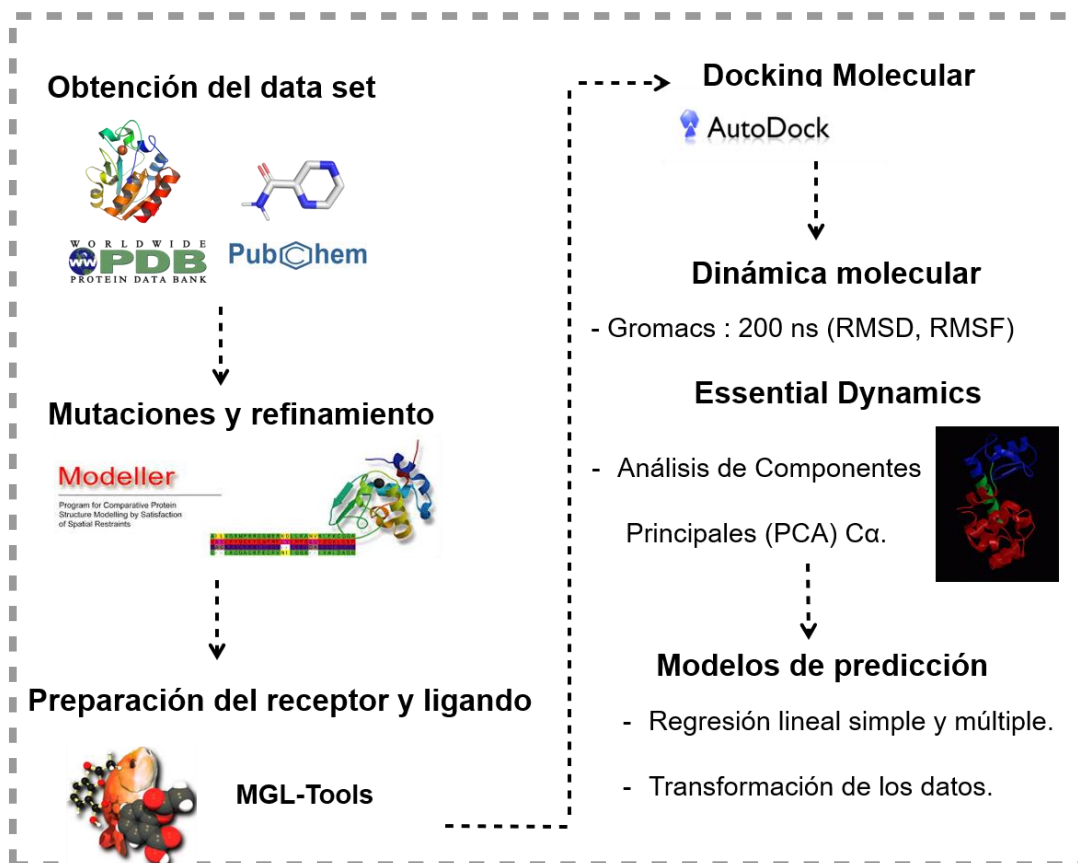
Se realizó un análisis de regresión lineal múltiple, en Epi Info v7.2, entre los parámetros geométricos, fisicoquímicos y los parámetros enzimáticos determinados experimentalmente tomando los datos del primer batch. Para la visualización de las líneas de tendencias, se realizó en un cuadernillo de código, Jupyter Notebook, que usa *Python scripting* (Anexo 4). Para ello, se utilizaron las librerías *Pandas*, *Numpy*, *Matplotlib*, *Sklearn* y *Pylab*, implementadas en *Python 3.7*.

Asimismo, se elaboró un modelo utilizando valores de estabilidad estimados con el RMSF de las dinámicas moleculares. En este modelo, se utilizó una transformación lineal logarítmica, descrita en la ecuación 4, para los parámetros cinéticos, debido a que estos tienden a seguir una distribución logarítmica-normal.

$$\overline{\log_{10}(x)} = \log_{10} \left( \frac{\bar{x}^2}{\sqrt{\bar{x}^2 + \sigma_x^2}} \right) \quad (4)$$

Donde x representan el valor del valor cinético y  $\sigma^2$  representa la varianza de los datos. Estos nuevos valores fueron ingresado al modelo haciendo uso de un *in-house R script*. Además, los intervalos de confianza (I.C.) fueron estimados al 95% para los coeficientes de regresión y el intercepto. Para el caso preciso del I.C. del  $r^2$  de cada modelo, se estimó mediante un re-muestreo (*bootstrapping*) de los datos usando un número de muestreo = 5000, en R-Studio.

Para tener un mejor panorama de los procedimientos bioinformáticos y de predicción realizados elaboramos un workflow (figura 3).



**Figura 3.** Workflow de los procedimientos bioinformáticos y de predicción.

### 3.15 Especificidad y sensibilidad

Debido a que contamos con parámetros cinéticos determinados experimentalmente para distintas mutaciones en *PZasa*, consideramos apropiado evaluar la sensibilidad y especificidad usando el modelo construido y los datos del segundo batch. Bajo estas circunstancias, tomamos la sensibilidad como la capacidad del modelo para clasificar correctamente una cepa resistente (ecuación 5), mientras que, la especificidad es la capacidad para clasificar correctamente una cepa sensible (ecuación 6).

$$\text{Sensibilidad} = VP / (VP + FN) \quad (5)$$

$$\text{Especificidad} = VN / (VN + FP) \quad (6)$$

Donde, VP son los verdaderos positivos (verdaderos resistentes), FN son los falsos negativos (falso sensibles), VN son los verdaderos negativos (verdaderos sensibles) y FP son los falsos positivos (falsos resistentes).

La especificidad y sensibilidad del modelo fueron evaluados tomando 14 datos de entradas, 7 provienen de mutaciones (que no fueron ingresadas en el modelo de regresión o batch 2) y 7 adicionales que fueron obtenidas de los componentes principales 1-7 del análisis de PCA. Para este análisis se utilizó el procedimiento recomendado (49).

### **3.16 Renderizado y ploteo**

Las imágenes fueron procesadas en PyMol y renderizadas a 150 dpi. Los plots fueron realizados usando el software Matplotlib (50).

## IV. RESULTADOS

### 4.1 Parámetros cinéticos

De acuerdo a lo reportado, 35 mutaciones puntuales en la enzima *PZasa* están asociadas de resistencia a pirazinamida (24). Para cada una de estas mutaciones, los parámetros cinéticos (enzimáticos) han sido descritos a nivel experimental; como la constante catalítica (*Kcat*), *Km*, eficiencia enzimática (*Eff*) y actividad enzimática (*Act.*) (Tabla 1). Dichos parámetros son expresados como porcentajes, correspondiendo a la actividad cinética del *wild type* el valor del 100%.

**Tabla 2.** Parámetros cinéticos de la *PZasa* de *M. tuberculosis wild type* y mutada.

Mutación	<i>Kcat</i>	<i>Kcat</i>	<i>Km</i>	<i>Km</i>	<i>Eff.</i>	<i>Eff.</i>	<i>Act.</i>	<i>Act.</i>	
	(min <sup>-1</sup> )	(%)	(mM)	(%)	(mM <sup>-1</sup> min <sup>-1</sup> )	(%)	(μmol POA min <sup>-1</sup> mg <sup>-1</sup> <i>PZsa</i> )	(%)	
WT*	1005.410	100.000	1.240	100.000	806.600	100.000	38.400	100.000	
A102V	1084.220	107.839	1.581	127.479	685.890	85.035	34.316	89.364	
C14G	25.110	2.497	1.327	107.014	18.920	2.346	0.902	2.350	
D12A*	245.170	24.385	0.990	79.839	248.790	30.844	9.240	24.063	
D12G*	368.460	36.648	0.550	44.355	716.830	88.871	14.000	36.458	
D136G	490.270	48.763	3.680	296.774	133.030	16.493	12.340	32.135	
D49N*	1.530	0.152	3.090	249.194	0.550	0.068	0.045	0.117	
Batch I	F58L	355.880	35.397	1.297	104.627	274.300	34.007	14.132	36.801
	F94L*	712.920	70.908	2.000	161.290	348.600	43.218	21.190	55.182
	G24D*	100.130	9.959	0.420	33.871	236.200	29.283	4.280	11.146
	G78C*	105.160	10.459	1.070	86.290	97.650	12.106	6.960	18.125
	H51R*	0.170	0.017	1.410	113.710	0.120	0.015	0.006	0.016
	H57R	12.020	1.196	1.099	88.609	10.940	1.356	0.505	1.316
	H71Y	69.840	6.946	10.657	859.464	9.320	1.155	1.425	3.710
	K48T*	241.820	24.052	0.440	35.484	551.200	68.336	10.450	27.214
	L116P*	1324.500	131.737	1.560	125.806	847.660	105.091	50.150	130.599
	L4S	0.001	-	0.201	16.173	0.003	0.000	0.015	0.039

P54L	141.195	14.044	1.001	80.763	140.989	17.479	5.529	14.398
P62L	203.859	20.276	0.779	62.802	261.782	32.455	10.129	26.378
Q10P	37.665	3.746	0.650	52.431	57.932	7.182	1.978	5.151
R29P	59.000	5.868	0.225	18.117	262.881	32.591	2.761	7.189
T135P*	0.450	0.045	0.930	75.000	-	-	0.020	0.052
T142A	0.040	0.004	1.130	91.090	0.040	0.005	0.002	0.005
T160K	3.500	0.348	13.630	1099.19	0.260	0.032	0.038	0.100
T76P*	202.980	20.189	0.310	25.000	650.330	80.626	8.990	23.411
V139A	628.710	62.533	25.422	2050.14	24.730	3.066	3.455	8.997
W119L	17.090	1.700	1.234	99.508	13.850	1.717	0.883	2.299
Y34D*	386.400	38.432	0.830	66.935	460.190	57.053	20.580	53.594
Y64D	1357.250	134.995	1.548	124.870	876.550	108.672	57.830	150.598
<b>Desv. Stdr</b>	394.114	39.523	5.350	431.454	286.567	35.528	14.797	38.534
<b>WT</b>	739.480	100.000	1.162	100.000	648.776	100.000	27.607	100.000
A171T	408.035	55.179	1.501	129.177	283.986	43.773	16.422	59.484
A46V	244.662	33.086	1.121	96.436	260.875	40.210	11.433	41.411
L172P	11.730	1.586	1.469	126.390	10.877	1.677	0.500	1.811
M175V	199.743	27.011	2.260	194.516	96.922	14.939	5.776	20.921
P62R	206.273	27.894	1.469	126.430	141.472	21.806	7.889	28.574
V125F	609.467	82.418	1.678	144.418	407.886	62.870	19.996	72.428
V180F	474.632	64.185	6.309	542.873	149.531	23.048	9.104	32.975
<b>Desv. Stdr</b>	201.046	27.187	1.819	156.540	132.920	20.488	6.537	23.678

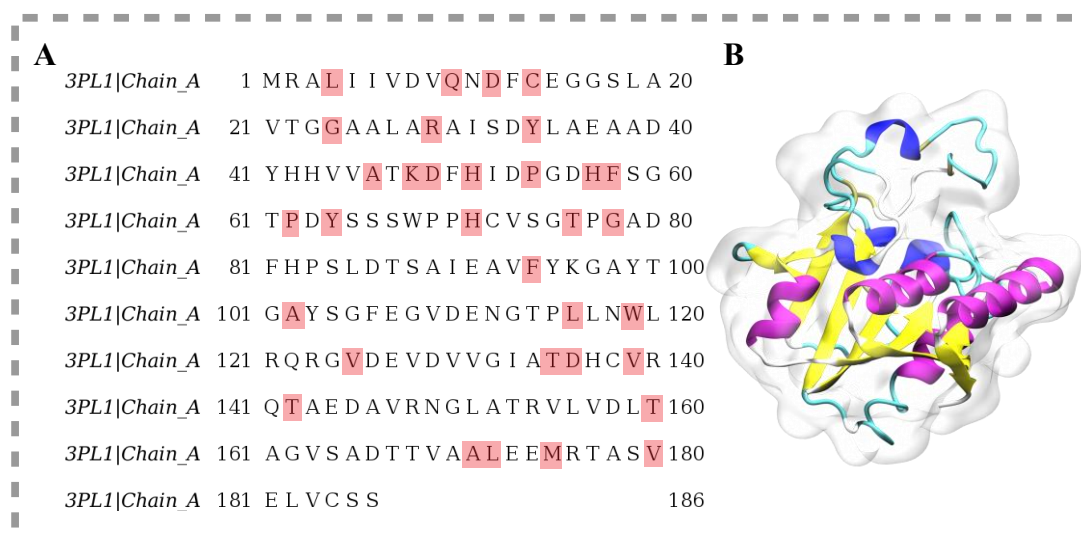
Los datos indicados con un \* han sido reportados previamente (24). Los demás datos han sido trabajados con el mismo protocolo en el laboratorio.

Tomando como referencia los valores estimados para la enzima *wild type* (WT), la menor variación en el *Kcat* (0.004%) se registró en T142A, para el *Km* (16.173%) en L4S, para la *Eff.* (0.005%) y *Act.* (0.005%) en T142A; mientras que, la mayor variación en el *Kcat* (135.00 %) se registró en Y64D, para el *Km* (2050.14%) en V139A, para la *Eff.* (108.67%) en Y64D y para la *Act.* (150.60%). Cabe destacar que las fluctuaciones son diversas, evidenciándose de esta forma el efecto de una única mutación puntual sobre la función de la enzima *PZasa* en *M. tuberculosis*.

Para los análisis de regresión, con los datos obtenidos en las simulaciones, solo se usaron los valores expresados en %.

#### 4.2 Descripción del receptor y las estructuras mutadas

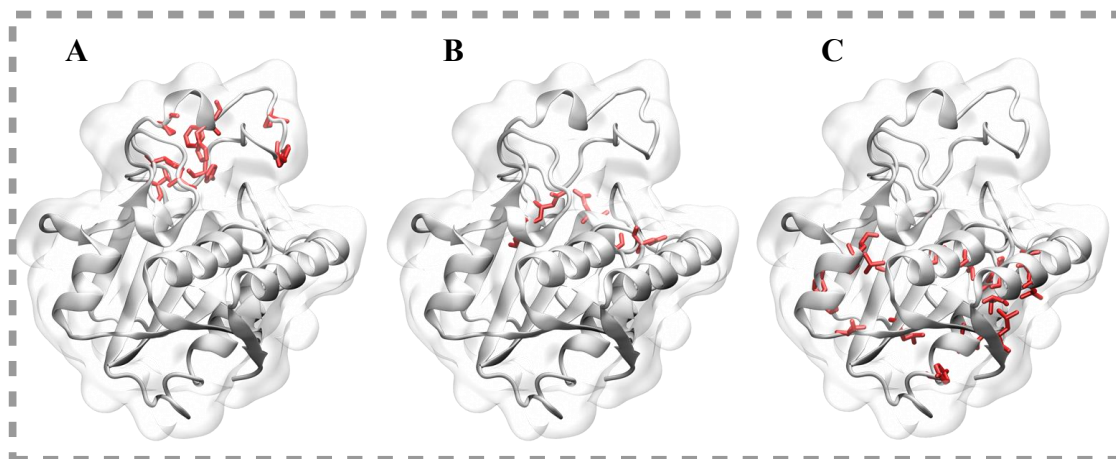
La enzima *PZasa* está compuesta por un monómero de 186 aminoácidos (Figura 3), con 2696 átomos (C<sub>856</sub>H<sub>1317</sub>N<sub>235</sub>O<sub>282</sub>S<sub>6</sub>). Tiene un peso molecular de 19.61 kilodatos (kDa) y un punto isoeléctrico (pI) de 4.43. El tiempo de vida medio estimado es de 30 horas, siendo >20 horas en levaduras y >10 horas en cultivos de *Escherichia coli*. El índice de inestabilidad (II) es de 10.35, siendo clasificada como una proteína estable. Además, el índice alifático (AI) es de 83.92, infiriéndose que es una proteína altamente termoestable (51).



**Figura 4.** A) Secuencia primaria de la enzima *PZasa* de *M. tuberculosis*. Las mutaciones reportadas están resaltadas en rojo. Algunos aminoácidos pueden sufrir más de una mutación (D12 y P62). B) Estructura tridimensional del cristal de la *PZasa* (58). Las hélices  $\alpha$  están coloreadas en morado y las láminas  $\beta$  en amarillo.

Para un mejor entendimiento de cómo se distribuyen las mutaciones en la estructura de la *PZasa*, estas se han organizado en 3 grupos principales (Figura 4). En el primer

grupo, se encuentran las mutaciones que ocurren dentro del sitio de coordinación metálico ( $\text{Fe}^+$ ), representado por las mutaciones en las posiciones A102V, F58L, H51R, H57R, H71Y, P54L, P62L, P62R, T76P y Y64D. Dichas mutaciones se distribuyen a lo largo de un *loop* (Figura 4A), cuya región ha sido determinada como la más flexible durante las simulaciones.



**Figura 5.** Distribución de las mutaciones puntuales caracterizadas en la enzima *PZasa* de *M. tuberculosis*. A) Mutaciones relacionadas con el centro de coordinación metálico. B) Mutaciones cercanas al sitio catalítico. C) Mutaciones en regiones distantes al sitio activo y de coordinación metálico.

En el segundo grupo se encuentran las mutaciones C14G, D12A, D12G, D136G, D49N, K48T y Q10P, las cuales componen el bolsillo (*pocket*) o sitio activo de la proteína (Figura 4B). En esta región se realiza la catálisis de la pirazinamida (PZ) en ácido pirazinóico (POA).

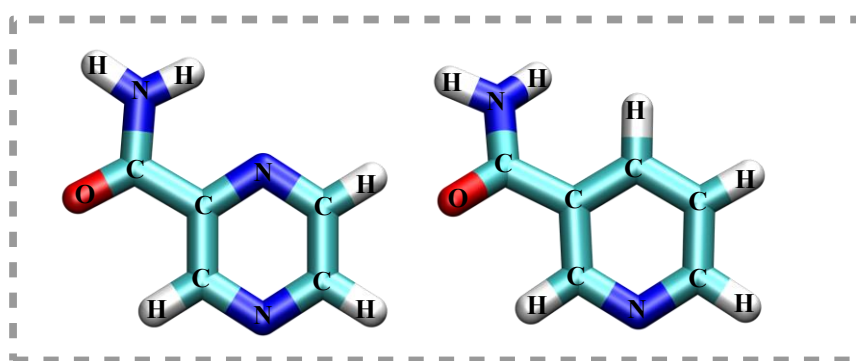
Por último, en el tercer grupo se incluyó a las mutaciones no cercanas a alguno de los sitios mencionados, como F94L, G24D, G78C, L116P, L4S, R29P, T135P, T142A, T160K, V139A, W119L, Y34D, A171T, A46V, L172P, M175V, V125F y V180F (Figura 4C). Si bien estas últimas mutaciones no están directamente

relacionadas con los aminoácidos que confieren la actividad, ellas pueden tener alguna actividad sobre la estabilidad y la dinámica de la enzima, induciendo modificaciones de largo alcance, evento que suele ser conocido como modificaciones alostéricas (52,53).

Los modelos de las estructuras mutadas realizadas con el software Modeller no difirieron significativamente en los valores de RMSD ( $< 0.0001$ ), debido a que el algoritmo de gradiente conjugada que se utilizó únicamente refina la posición del aminoácido mutado. Dicha limitación fue abordada posteriormente con los análisis de simulación de dinámica molecular, en donde se observó la influencia de las mutaciones en la conformación y dinámica de la estructura.

#### 4.3 Descripción del ligando

La pirazinamida (PZ) es una molécula derivada de la nicotinamida (Figura 5). Está compuesta de 14 átomos ( $C_5H_5N_3O$ ) y tiene un peso molecular de  $123.11 \text{ g mol}^{-1}$ . Su tiempo de vida media es de 9 a 10 horas en pacientes con metabolismo normal y es activa a pH 5.6 (54).



**Figura 6.** Estructura molecular de la pirazinamida (izquierda) y nicotinamina (derecha).

#### 4.4 Docking molecular

Los análisis de docking molecular fueron realizados con las estructuras mutadas y refinadas en el paso anterior. Este análisis permitió comparar los modos de acoplamiento entre la droga PZ y las diferentes estructuras mutadas de la *PZasa*. Los resultados de estos cálculos demuestran una ligera variación entre las metodologías abordadas, encontrándose energías más favorables en los dockings del tipo hidratado.

**Tabla 3.** Diferentes enfoques de docking molecular en el acoplamiento ligando-receptor.

Mutación	<i>Binding energy (Kcal mol<sup>-1</sup>)</i>				
	Sin ion Fe <sup>+</sup>		Con ion Fe <sup>+</sup>		
	Hidratado	No hidratado	Hidratado	No hidratado (Fe <sup>+</sup> 2.0)	No hidratado (Fe <sup>+</sup> 0.0)
A102V	-6.55	-4.46	-6.52	-4.84	-4.51
C14G	-6.48	-4.46	-6.39	-5.52	-4.46
D12A	-6.49	-4.46	-6.31	-5.57	-4.46
D12G	-6.52	-4.46	-6.31	-5.57	-4.46
D136G	-6.49	-4.45	-6.32	-5.57	-4.45
D49N	-6.55	-4.48	-6.31	-5.57	-4.48
F58L	-6.63	-4.48	-6.50	-6.10	-4.51
F94L	-6.49	-4.46	-6.31	-5.57	-4.46
G24D	-6.54	-4.46	-6.31	-5.57	-4.46
G78C	-6.55	-4.45	-6.31	-5.56	-4.45
H51R	-6.50	-4.46	-6.32	-5.57	-4.46
H57R	-6.27	-4.46	-6.28	-5.37	-4.46
H71Y	-6.60	-4.60	-6.53	-5.25	-4.46
K48T	-6.49	-4.60	-6.45	-4.52	-4.62
L116P	-6.52	-4.46	-6.30	-5.57	-4.46
L4S	-6.49	-4.46	-6.32	-5.56	-4.47
P54L	-6.55	-4.46	-6.39	-5.56	-4.46

P62L	-6.50	-4.46	-6.32	-5.57	-4.46
Q10P	-6.52	-4.45	-6.31	-5.56	-4.46
R29P		-4.47	-6.31	-5.56	-4.46
T135P	-6.55	-4.46	-6.31	-5.57	-4.46
T142A	-6.57	-4.47	-6.33	-5.57	-4.47
T160K	-6.52	-4.47	-6.32	-5.57	-4.47
T76P	-6.48	-4.46	-6.31	-5.56	-4.46
V139A	-6.49	-4.46	-6.31	-5.57	-4.46
W119L	-6.55	-4.45	-6.30	-5.57	-4.45
Y34D	-6.51	-4.45	-6.31	-5.57	-4.46
Y64D	-6.50	-4.46	-6.31	-5.57	-4.46
<b>WT</b>	-6.55	-4.45	-6.32	-5.55	-4.44
A171T	-6.55	-4.46	-6.31	-5.57	-4.46
A46V	-6.55	-4.46	-6.32	-5.57	-4.46
L172P	-6.48	-4.46	-6.32	-5.57	-4.46
M175V	-6.55	-4.46	-6.31	-5.56	-4.46
P62R	-6.55	-4.47	-6.30	-5.65	-4.47
V125F	-6.50	-4.46	-6.31	-5.57	-4.46
V180F	-6.50	-4.46	-6.30	-5.56	-4.46
<b>Desv. Stdr</b>	$\pm 0.055$	$\pm 0.033$	$\pm 0.063$	$\pm 0.238$	$\pm 0.028$

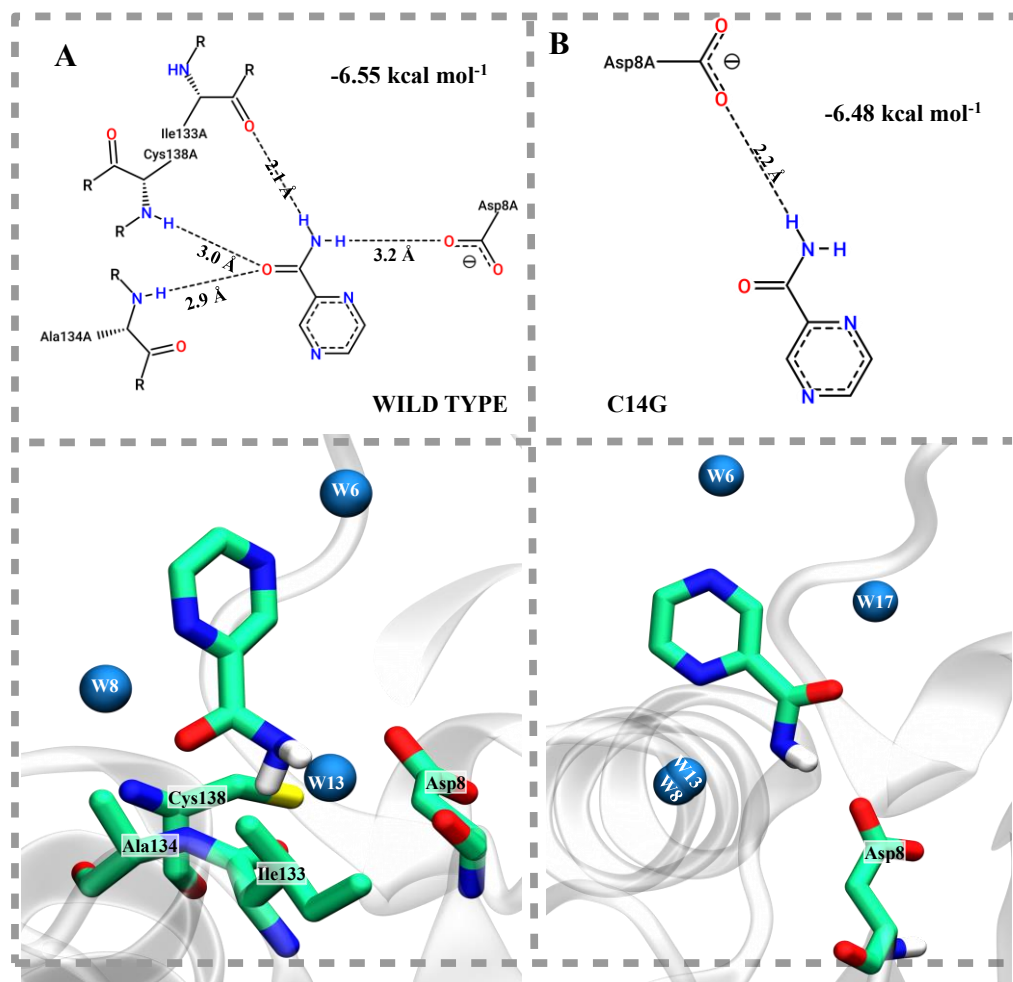
La energía de interacción más favorable (menor energía) para la estructura *PZasa wild type* ( $-6.55 \text{ kcal mol}^{-1}$ ) fue obtenida en el docking del tipo hidratado sin ion  $\text{Fe}^+$ . Tomando este valor como un *cut-off*, se observan que las energías fluctúan con las mutaciones incrementándose y en otras disminuyendo. El mismo patrón fue observado en las diferentes metodologías empleadas (Tabla 3). Por otro lado, considerando las energías de interacción en los acoplamientos con estructuras mutadas que ocurren únicamente en el sitio activo de la *PZasa*, notamos que estas no difieren significativamente entre sí. La mayor energía de interacción para este sitio se obtuvo con la mutación C14G ( $-6.48 \text{ kcal mol}^{-1}$ ), mientras que la menor energía en A102V y D49N ( $-6.55 \text{ kcal mol}^{-1}$  para ambas).

Cabe señalar que, dentro de cada metodología de docking molecular desarrollada, las energías de interacción presentaron una desviación estándar muy ligera (Tabla 3). La máxima desviación estándar ( $\pm 0.238 \text{ kcal mol}^{-1}$ ) se obtuvo con el docking del tipo no hidratado en presencia de ion metálico, mientras que la mínima ( $\pm 0.028 \text{ kcal mol}^{-1}$ ) con el docking no hidratado con ion metálico y carga  $+2.0$ .

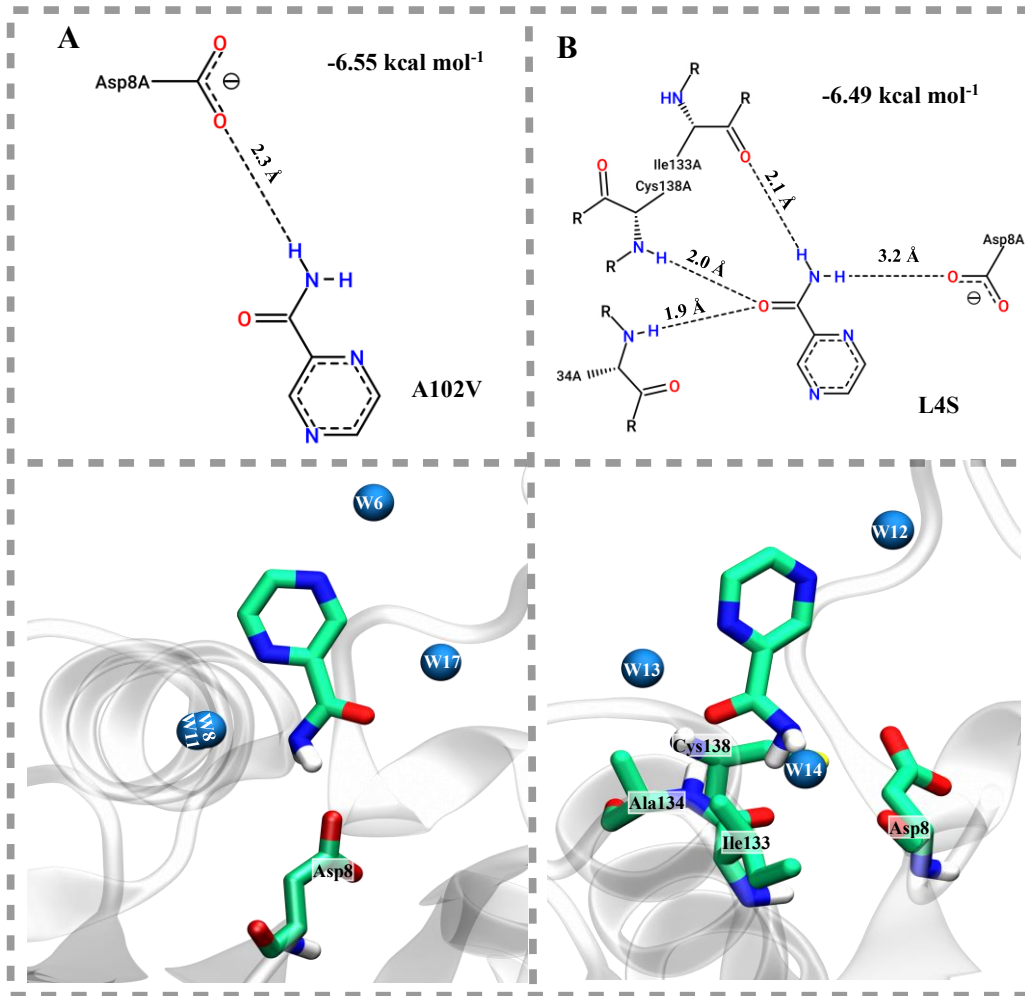
Nuestros análisis de interacciones permitieron identificar los modos de acoplamiento entre la pirazinamidasa y la *PZasa*. En el análisis con la menor energía de interacción ( $-6.55 \text{ kcal mol}^{-1}$ ), se observó que la droga se acopla a la proteína formando 4 puentes de hidrógeno con 4 residuos localizados en el *binding site*, Asp8, Ile133, Ala134 y Cys138. Las distancias entre los grupos químicos de la droga (carbonilo y amino) y los residuos de la enzima son de 3.2, 2.1, 2.9 y 3.0 Å, respectivamente. Asimismo, el docking hidratado indicó que 3 moléculas de agua contribuyen significativamente con la interacción. Dichas moléculas se encuentran alrededor de 2.0 y 3.0 Å de la droga. Cabe destacar que, el modo de acoplamiento determinado en nuestras simulaciones es altamente similar al reportado en el cristal (41); además, se observó que nuestras perdiciones concuerdan con los aminoácidos catalíticos reportados y una de las moléculas de agua (W13) se posiciona similarmente a una de las moléculas de agua que confiere una función estabilizadora, cercana al residuo Asp8 (Figura 6A).

Por otro lado, en el análisis del acoplamiento de la droga con la mutación que presentó la mayor energía de interacción ( $-6.48 \text{ kcal mol}^{-1}$ ), se determinó que esta interacciona únicamente con el residuo Asp8, formando un puente de hidrógeno a una distancia de 2.2 Å. Interesantemente, la mutación ocasionó la pérdida de la interacción con los residuos antes mencionados (Ile133, Ala134 y Cys138), sin

embargo, se ha observado que 4 moléculas de agua presentes (W6, W8, W13 y W17) favorecen la interacción dentro del sitio activo (Figura 6B).



**Figura 7.** Interacciones moleculares entre la pirazinamida y la enzima *PZasa* (superior). Modos de acoplamiento (inferior), en donde las moléculas de agua (W) están representadas como esferas. A) Acoplamiento del ligando hidratado en el sitio catalítico de la estructura *wild type*. B) Acoplamiento del ligando hidratado en el sitio catalítico de la estructura mutada C14G.

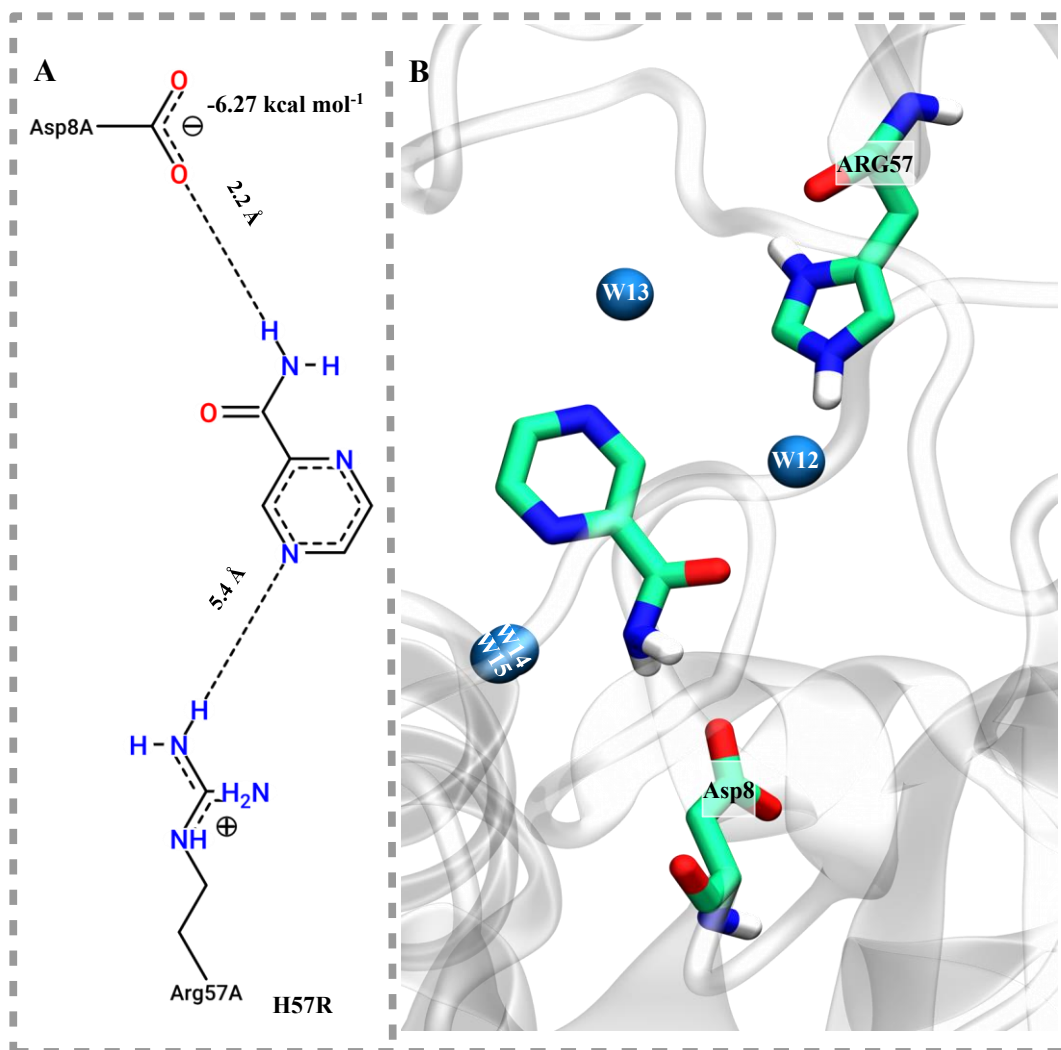


**Figura 8.** Interacciones moleculares entre la pirazinamida y la enzima *PZasa* (superior). Modos de acoplamiento (inferior), en donde las moléculas de agua (W) están representadas como esferas. A) Acoplamiento del ligando hidratado en el sitio catalítico de la estructura *wild type*. B) Acoplamiento del ligando hidratado en el sitio catalítico de la estructura mutada L4S.

En la estructura mutada A102V, se observó una interacción con el residuo Asp8, a una distancia de 2.3 Å. Similarmente, las interacciones perdidas son reemplazadas por contactos favorecidos por moléculas de agua (W6, W8, W11 y W17) (Figura 7A). Mientras que en la estructura mutada L4S, las interacciones son similares a las

obtenidas en el modelo *wild type*, con el mismo número de moléculas de agua (Figura 7B).

Particularmente, en el análisis de acoplamiento de la droga con la estructura mutada D49N no se observó ninguna interacción con los residuos catalíticos. Sin embargo, se determinó que las moléculas de agua favorecen el contacto en el sitio activo de este acoplamiento. Asimismo, cabe resaltar el modo acoplamiento de la droga con la estructura mutada H57R, en la cual se presentó una interacción con este residuo mutado, favorecida por 4 moléculas de agua, similarmente (Figura 8). Cabe mencionar que, en todos los análisis de acoplamiento la interacción con el residuo Asp8 se encuentra presente, tanto en el modelo *wild type* y mutantes, con excepción de D49N.

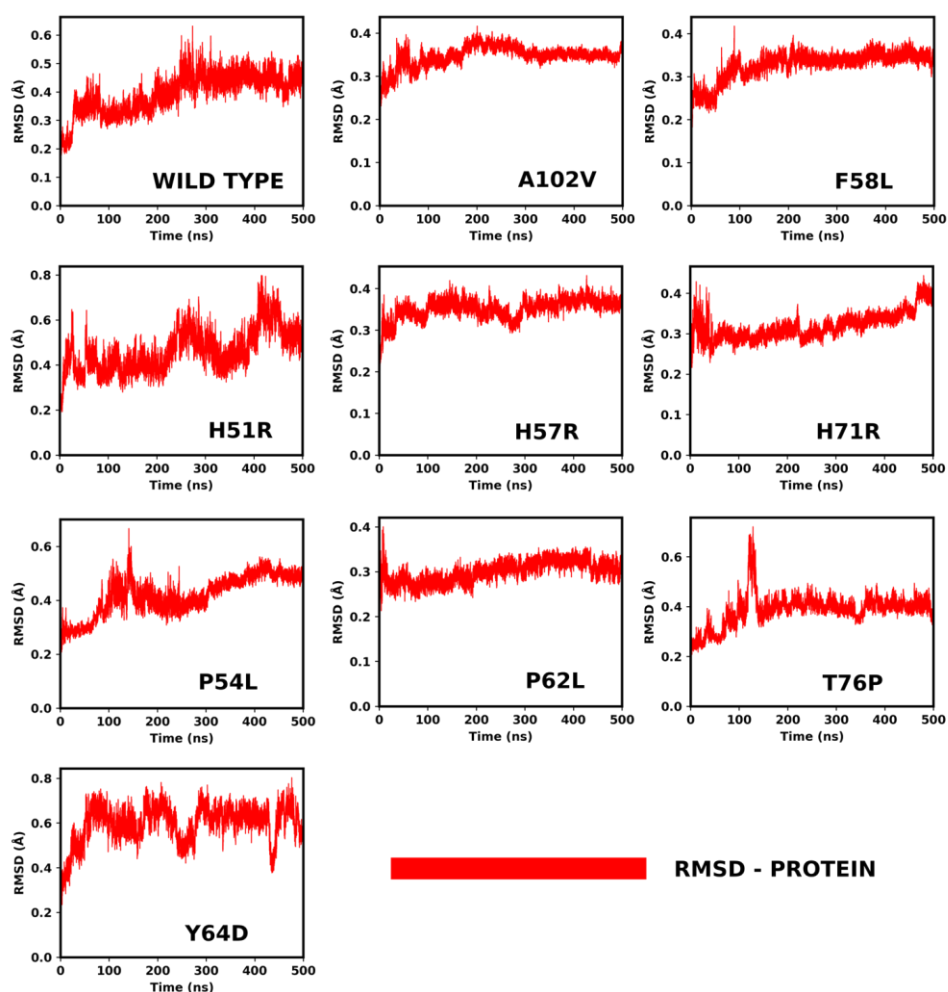


**Figura 9.** Interacciones moleculares entre la pirazinamida y la enzima *PZasa* mutada en H57R (izquierda). Modos de acoplamiento de la droga y las moléculas de agua en el sitio activo (derecha).

#### 4.5 Dinámica molecular

Mediante los análisis de dinámica molecular se ha determinado la influencia de cada mutación puntual sobre la dinámica de la *PZasa*. En las estructuras con mutaciones cercanas al centro de coordinación metálico, los valores de RMD fueron inferiores a 0.6 nm (6 Å), valores que convergen a partir de los 200 ns en el modelo *wild type*, mientras que la estructura mutada A46V converge luego de los ~300 ns.

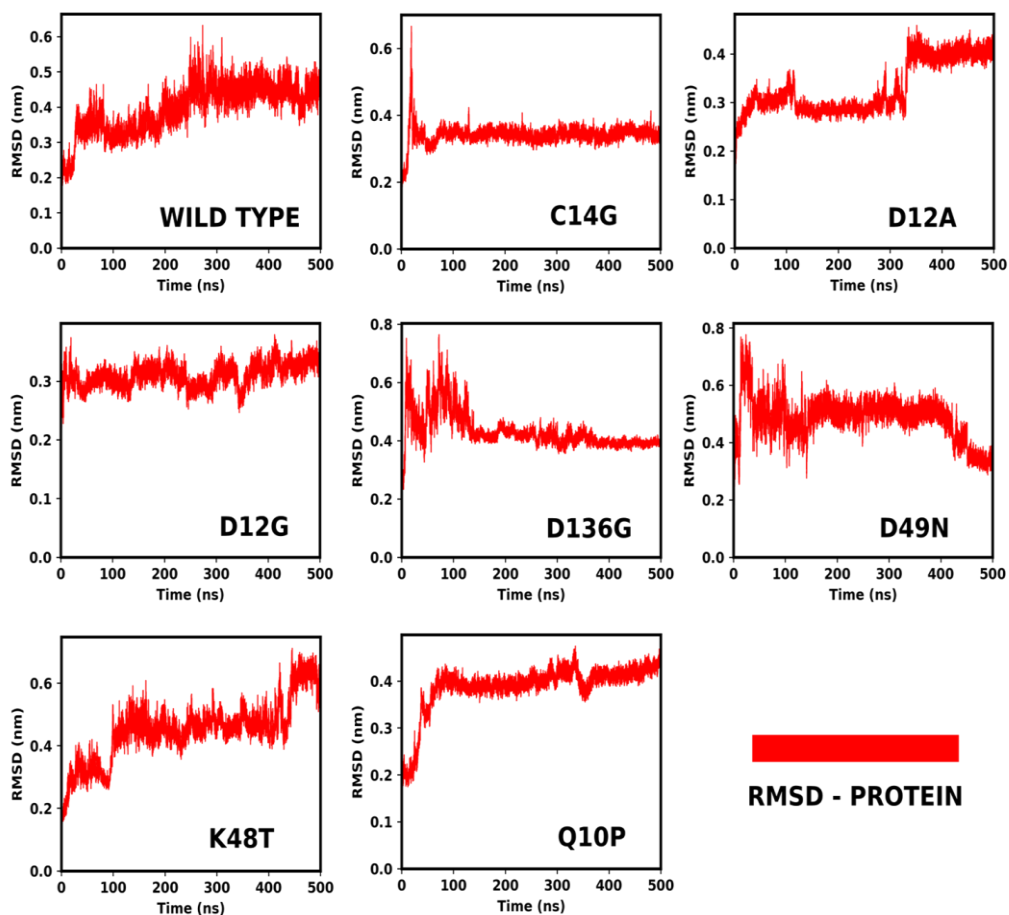
Las estructuras mutadas que pueden ser descritas como más estables, debido a su rápida convergencia, son F58L, P62L y T76P. Las estructuras que presentaron mayor inestabilidad fueron H51R y Y64D, fluctuaciones que han podido ser observadas y asociadas al loop en donde coordina el cofactor metálico de la enzima (Figura 9).



**Figura 10.** Fluctuación del RMSD de la enzima *PZasa wild type* y con mutaciones cercanas al sitio de coordinación metálico.

En relación a las mutaciones que ocurren cercanamente al sitio catalítico, los valores de RMSD presentaron fluctuaciones ligeramente superiores (hasta 0.8 nm). En estas mutaciones, únicamente C14G, D136G y Q10P alcanzan la convergencia

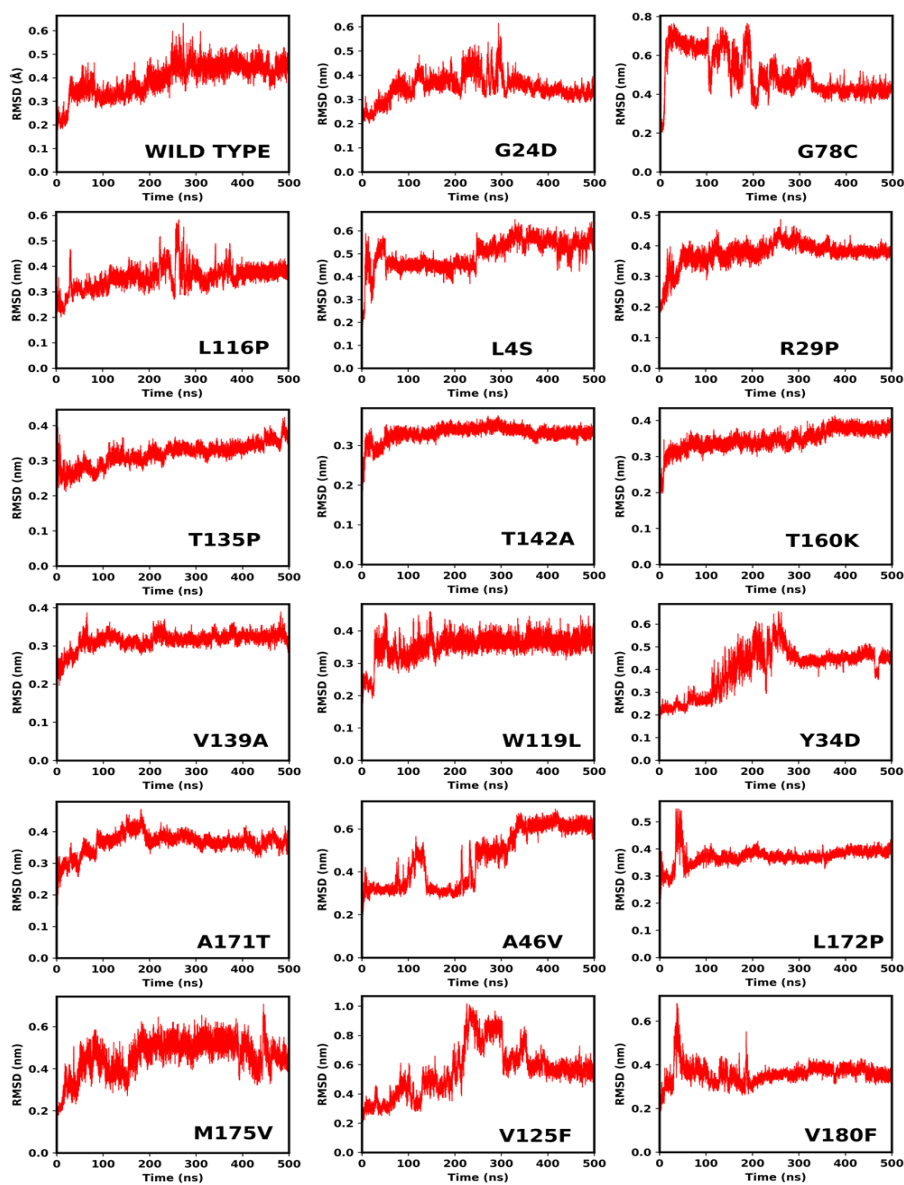
posterior a los 200 ns aproximadamente. Las mayores fluctuaciones de la estructura han sido determinadas en K48T. Asimismo, la mutación D49N causa fluctuaciones significativas que evitan que esta estructura converja a pesar del extenso tiempo de simulación (500 ns).



**Figura 11.** Fluctuación del RMSD de la enzima *PZasa wild type* y con mutaciones cercanas al sitio catalítico.

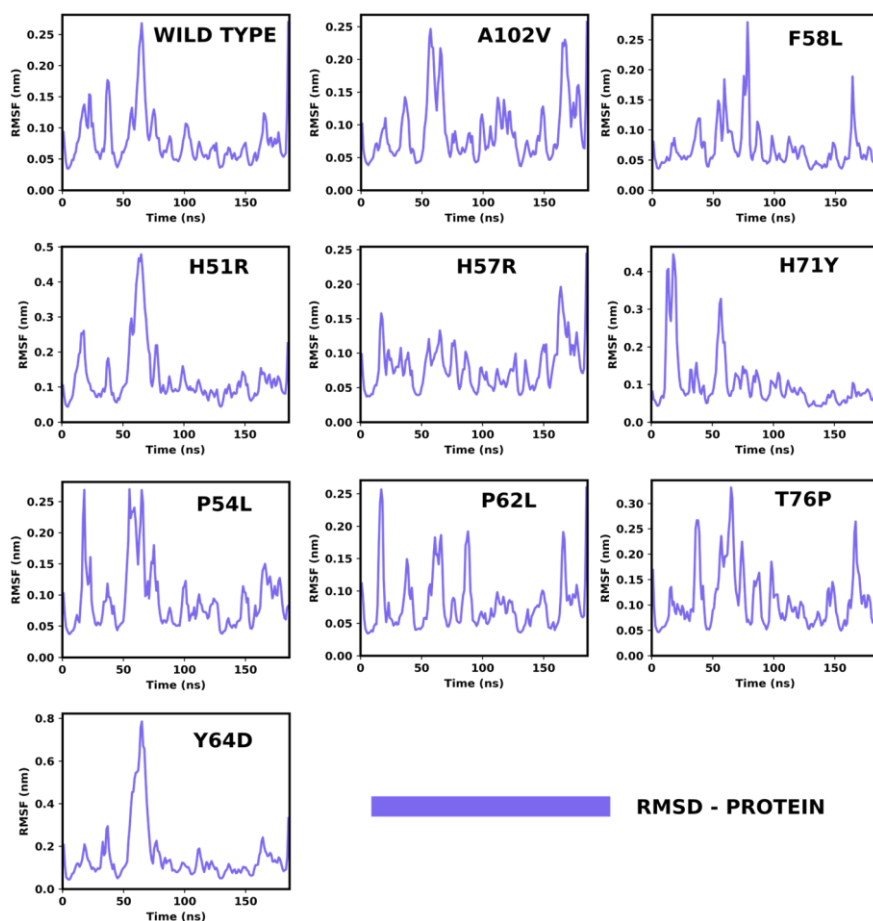
Asimismo, se ha determinado que ciertas mutaciones lejanas al sitio activo y de coordinación metálico ocasionan modificaciones estructurales significativas durante la dinámica de la proteína. La mutación G78C presentó un RMSD de hasta  $\sim 0.7$  nm y logra converger después de los  $\sim 350$  ns. Similarmente, G24D, Y34D, A46V y V125F convergen después de los  $\sim 200$  ns. Por otra parte, las mutaciones

R29P, T142A, T160K, V139A, W119L y L172P, presentaron fluctuaciones cercanas a 0.3 nm (3 Å), y estas convergen antes de los 100 ns. Mientras que, las mutaciones T135P y V125F presentan las mayores fluctuaciones, incluso hasta antes del término de la simulación.



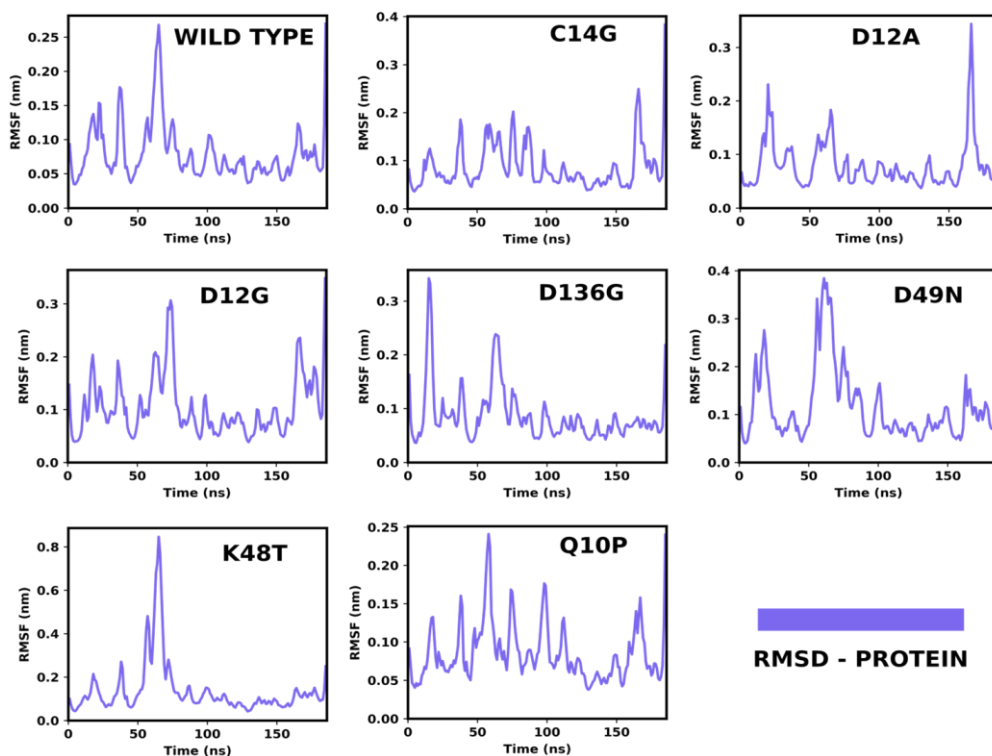
**Figura 12.** Fluctuación del RMSD de la enzima *PZasa wild type* y con mutaciones lejanas al sitio activo y de coordinación metálico.

Por otro lado, las fluctuaciones de los valores estimados del RMSF fueron más significativas, comparado con los valores del RMSD, y han sido determinadas para cada residuo en particular. La estructura *wild type* presenta la mayor fluctuación de RMSF a lo largo de los residuos Pro62, Asp63, Tyr64 y Ser65, la que es propia de la región de loop que coordina con el co-factor metálico. Similarmente, las estructuras mutadas A102V, H51R, P54L, T76P y Y64D tienen las mayores fluctuaciones en los mismos residuos. Cabe resaltar que, la mutación Y64D causó el mayor valor de RMSD, principalmente en el residuo Ser65 (0.79 nm) (Figura 12).



**Figura 13.** Fluctuaciones del RMSF de la enzima *PZasa wild type* y con mutaciones cercanas al sitio de coordinación metálico.

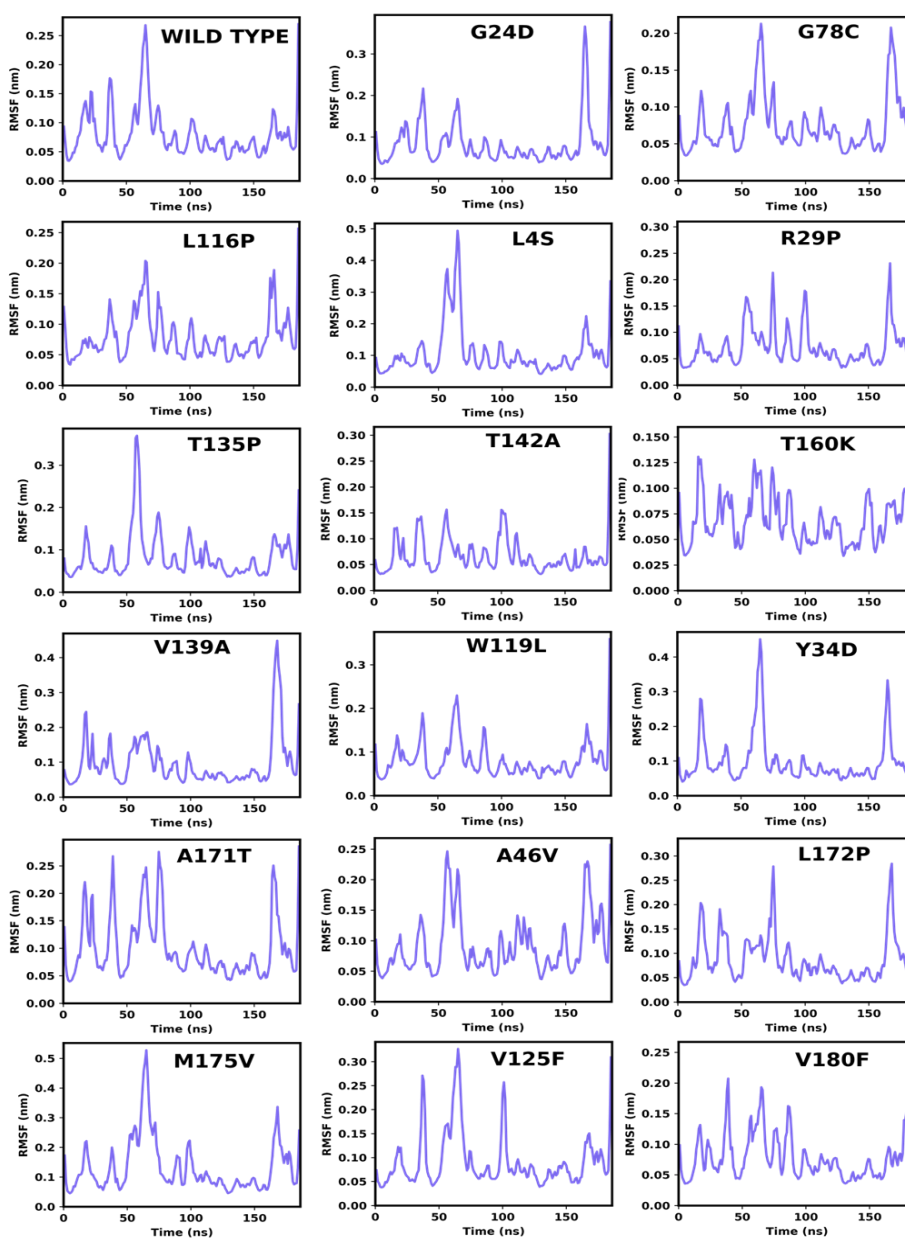
Por su parte, la estructura mutada A46V presentó una fluctuación similarmente alta en los residuos Ala165, Asp166, Thr167 y Thr168 ( $\sim 0.23$  nm). La mutación P54L ocasiona un valor significativamente alto en Ser18 (0.27 nm). Contrariamente, H57R no presenta la fluctuación característica del loop mencionado, en donde los valores del RMSF en los principales residuos de esta región son de  $\sim 0.11$  nm (Figura 12). Este resultado es concordante con lo obtenido en las simulaciones de acoplamiento, en donde la mutación favorece una interacción con la droga.



**Figura 14.** Fluctuaciones del RMSF de la enzima *PZasa wild type* y con mutaciones en el sitio catalítico.

En cuanto a las mutaciones en el sitio activo, K48T causó la mayor fluctuación de RMSF (0.85 nm) en el residuo Ser65. C14G causó fluctuaciones de  $\sim 0.22$  nm en los residuos próximos a Ala165, Asp166 y Thr167. Por su parte, la mutación D12G incrementa el RMSF en los residuos Cys72, Val73, Ser74, Gly75, en  $\sim 0.31$  nm.

Mientras que, D49N eleva los valores en los residuos Asp56, Gly60, Thr61, Pro62, Asp63, Tyr64, Ser65 y Ser66, en promedio de 3.70 nm. Asimismo, D136G presentó fluctuaciones mayores en los residuos Cys14, Glu15, Gly16 y Gly17, con un RMSF alrededor de 0.33 nm (Figura 13).



**Figura 15.** Fluctuación del RMSF de la enzima *PZasa wild type* y con mutaciones lejanas al sitio activo y de coordinación metálico.

En cuanto al RMSF en los sitios distantes de la cavidad catalítica y el sitio de coordinación metálico, se determinó que L4S genera la mayor fluctuación en Tyr64, Ser65 y Ser66, de hasta 0.49 nm. Otras mutaciones con RMSD significativo fueron G24D, T135P, L172P, M175V y V125F. Por su parte, la mutación T160K decrementa el RMSF, ocasionando una compactación en la estructura. Además, en todas se ha observado un alto valor de RMSF en el residuo Ser186, sin embargo, no consideramos este residuo como relevante debido a su localización en el extremo C-terminal de la proteína (Figura 14).

#### **4.6 Análisis de componentes principales (PCA – *essential dynamics*)**

El análisis de componentes principales permitió identificar las conformaciones con las mayores fluctuaciones (a nivel de conformación) durante la simulación. A partir de esta información fue posible obtener parámetros geométricos y fisicoquímicos para ser relacionados con la función biológica de la enzima (55).

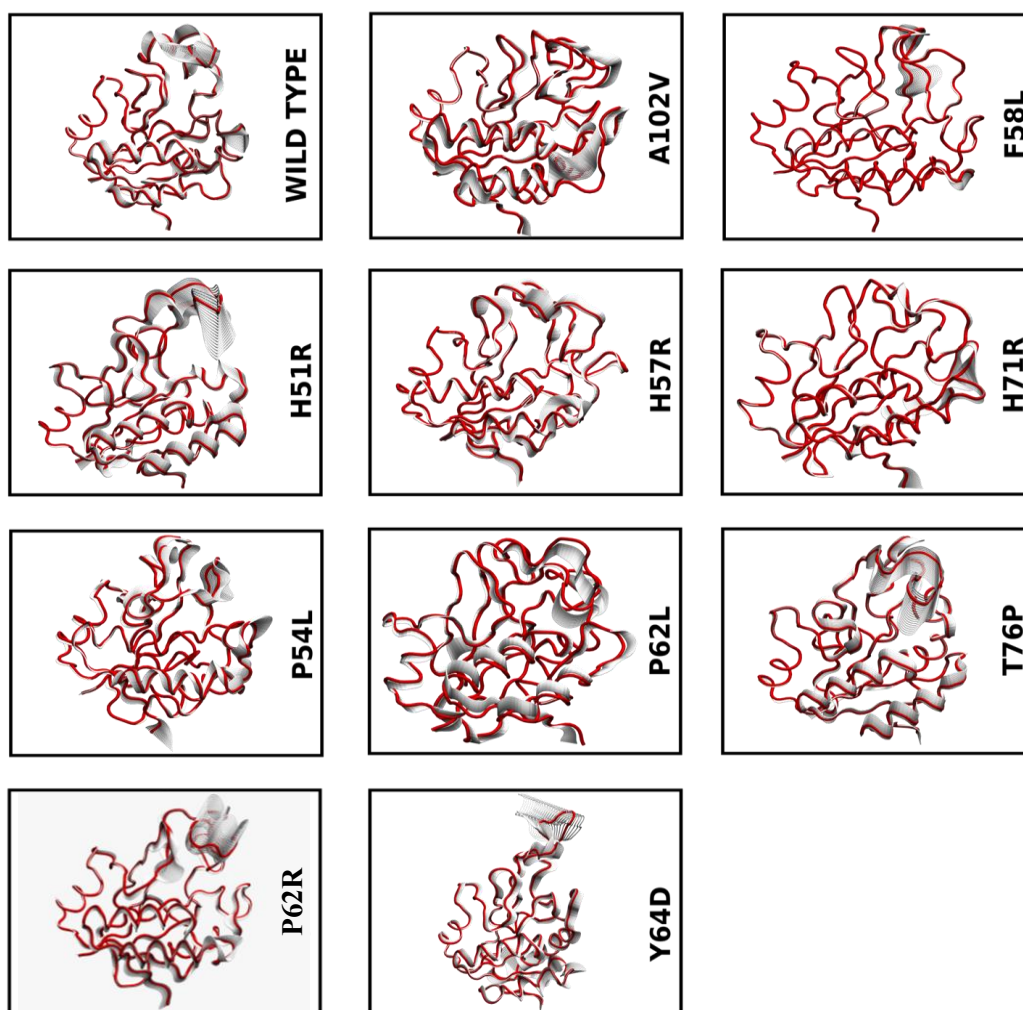
De los componentes principales (30) de la estructura *wild type*, se determinó que la región con mayor flexibilidad corresponde al *loop* en donde coordina el co-factor metálico. Además de ello, se observó otra región de *loop* que presenta gran flexibilidad, correspondiente a los residuos Val73, Ser74, Gly75, 76, Pro77, ThrGly78 y Ala79. Este último conecta la región del sitio de coordinación metálico con el sitio catalítico (Figura 15).

En general, podemos afirmar que todas las mutaciones modifican el comportamiento de la estructura. Sin embargo, dentro de estas mutaciones se destacan dos grupos principales, aquellas mutaciones que incrementan la flexibilidad (relajan la estructura) y aquellas que decrementan la flexibilidad (compactan la estructura).

En el grupo de las mutaciones que incrementan la flexibilidad se encuentran H51R, T76P, Y64D, en donde la última ocasionó la mayor fluctuación en la dinámica molecular de la enzima. Por su parte, A102V, F58L H57R, H71R, P54L y P62L compactan la estructura, ocasionando el cierre de la cavidad del sitio catalítico. Cabe destacar que estas mutaciones corresponden únicamente a la región que coordina con el co-factor metálico, en tal sentido nuestras observaciones confirmaron que las mutaciones que ocurren en esta región solo tienen influencia en esta misma (Figura 15).

Un patrón similar ha sido observado en el grupo de las mutaciones cercanas al sitio catalítico. En su mayoría, estas mutaciones ocasionan el cierre de la estructura, dentro de las cuales destacan a C14G, D12A, D12G y Q10P. Las mutaciones D136G, D49N y k48T aperturan ampliamente la cavidad. Asimismo, se ha observado que las regiones distantes mantienen una relativa estabilidad similar a la estructura *wild type* (Figura 16).

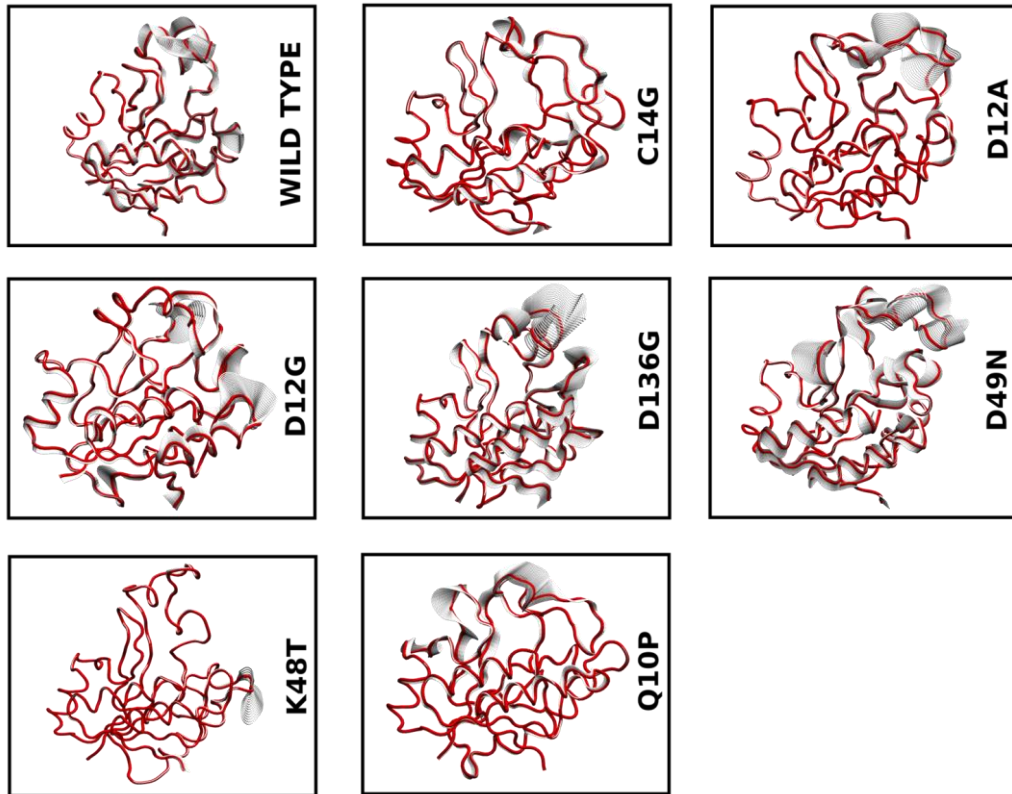
Asimismo, en este último grupo destacan las mutaciones V139A y Y34D. La primera ocasiona, en particular, el cierre de la cavidad que es sobrelapado por el loop del sitio de coordinación, pero, además, se ha observado una alta vibración de los residuos comprendidos entre la posición 166 y 176, los cuales conforman una estructura de hélice alfa que es cercana al sitio activo. Por su parte, Y34D ocasionó una modificación estructural en dichos residuos, haciendo que esta estructura secundaria se alargue (Figura 17).



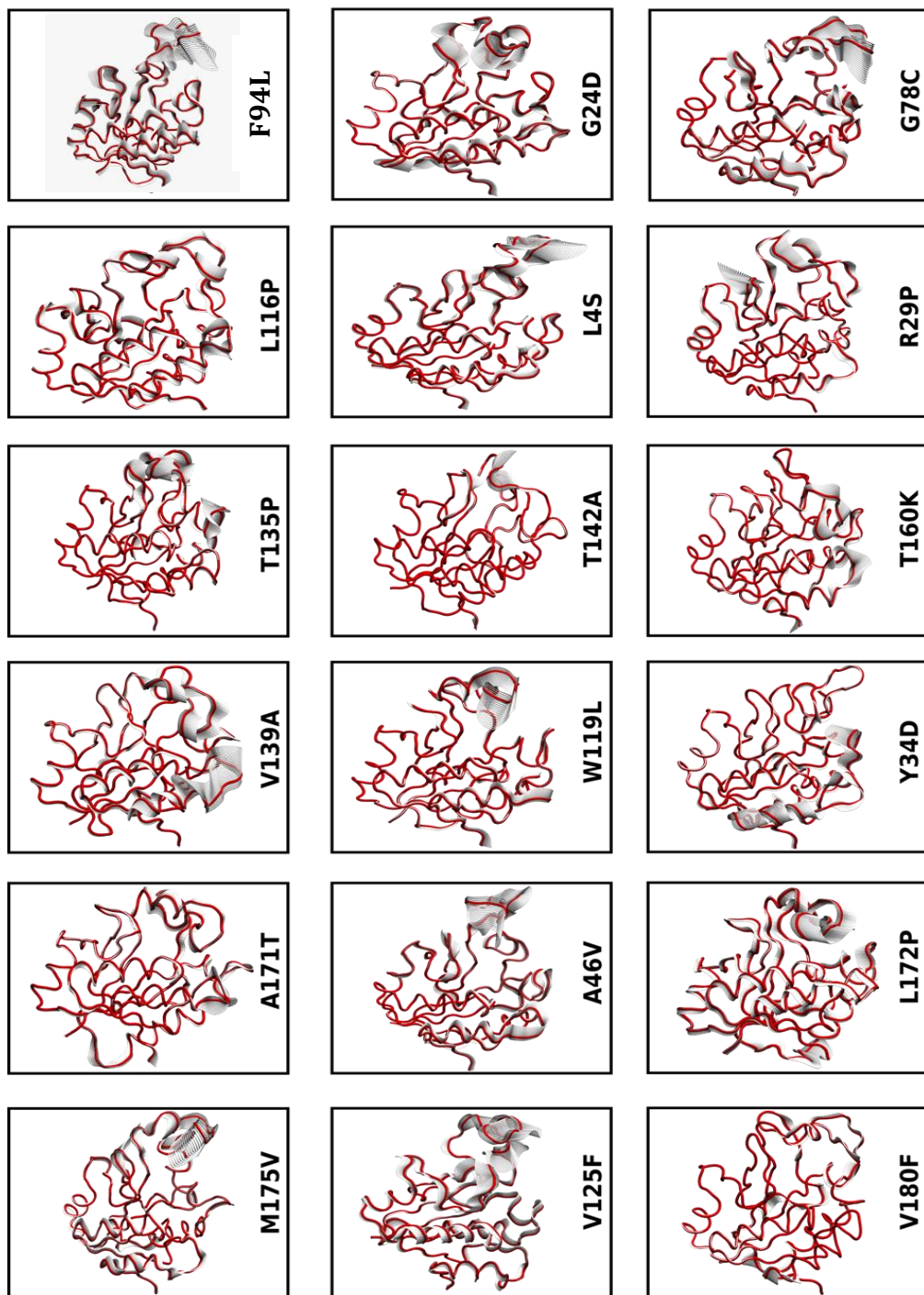
**Figura 16.** Análisis de componentes principales de la enzima *PZasa wild type* y con mutaciones cercanas al sitio de coordinación metálico. El promedio de las mayores fluctuaciones está representado en líneas rojas, mientras que las fluctuaciones extremas están como líneas blancas.

En tanto, las mutaciones que ocurren en sitios distantes presentaron un patrón similar, sin embargo, se ha podido observar que el efecto se da a largo alcance. Nuestras observaciones sugieren que L4S, G78C, A46V, M175V y V125F causan fluctuaciones en la región del *loop* de coordinación, mientras que R29P, T135P, T142A, T160K, A171T, y V180F causan el cierre del sitio catalítico. Es preciso

señalar que estas mutaciones están a una distancia de entre 16 y 24 Å, por lo que resulta evidente pensar en modificaciones estructurales alostéricas.



**Figura 17.** Análisis de componentes principales de la enzima *PZasa wild type* y con mutaciones en el sitio activo. El promedio de las mayores fluctuaciones está representado en líneas rojas, mientras que las fluctuaciones extremas están como líneas blancas.



**Figura 18.** Análisis de componentes principales de la enzima *PZasa wild type* y con mutaciones en el sitio activo. El promedio de las mayores fluctuaciones está representado en líneas rojas, mientras que las fluctuaciones extremas están como líneas blancas.

#### 4.7 Parámetros estructurales y fisicoquímicos

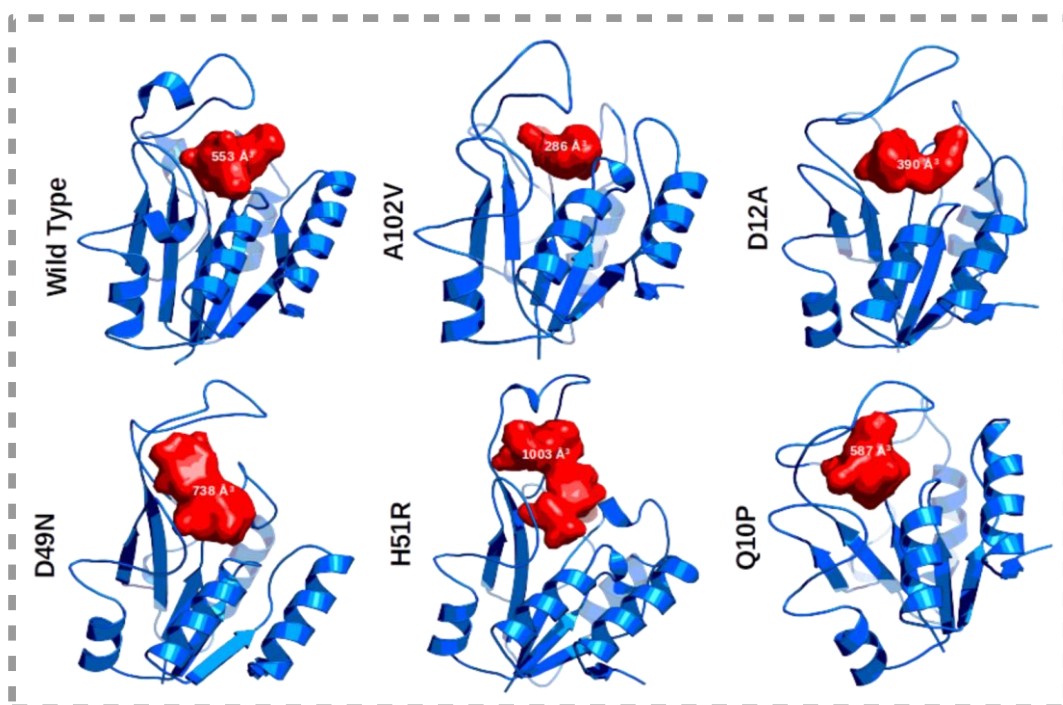
Los parámetros estructurales y fisicoquímicos fueron determinados, tanto para el modelo *wild type* y mutado, tomando la estructura promedio de las conformaciones obtenidas en el PCA. De los parámetros estudiados, las mayores varianzas fueron identificadas en los volúmenes que componen la cavidad del sitio activo, en donde se obtuvo una desviación estándar de  $\pm 371.90 \text{ \AA}^3$  entre las estructuras mutadas. La mutación L4S presentó la mayor fluctuación del volumen ( $1536 \text{ \AA}^3$ ), mientras que los volúmenes en P54L y P62L no pudieron ser determinados debido al cierre de la cavidad (Tabla 4).

Con respecto a los valores de distancia entre los residuos 57 y 49, la varianza en entre las estructuras mutadas fue de  $3.91 \text{ \AA}$ . La mutación K48T presentó la máxima distancia de  $19.60 \text{ \AA}$ , mientras que la mínima fue identificada en Y34D, la cual es la única que presenta un valor inferior a la distancia determinada en la estructura *wild type*. Por otro lado, los valores de RMSD fueron más altos en F58L, T135P, T160K, T76P, W119L y A171T, los cuales se localizan fuera del sitio activo.

**Tabla 4.** Parámetros estructurales de la *PZasa wild type* y mutada de *M. tuberculosis*. Los datos son obtenidos de las estructuras analizadas en el PCA.

Mutación	Vol. $\text{\AA}^3$	Dist. 57-49	RMSD (nm)	RMSF (residuo)								Rgx
				8	49	51	57	71	133	137	138	
WT	553	4.50	0.00	0.04	0.06	0.09	0.17	0.22	0.05	0.05	0.05	1.61
A102V	286	14.30	1.99	0.02	0.02	0.02	0.07	0.01	0.04	0.03	0.04	1.52
C14G	694	10.80	1.59	0.01	0.03	0.03	0.06	0.01	0.02	0.02	0.03	1.57
D12A	390	5.10	1.43	0.05	0.01	0.03	0.23	0.07	0.02	0.05	0.04	1.56
D12G	249	11.50	1.69	0.02	0.03	0.03	0.03	0.07	0.02	0.04	0.04	1.54

<b>D136G</b>	983	13.50	1.99	0.1	0.05	0.08	0.19	0.15	0.03	0.08	0.05	1.57
<b>D49N</b>	738	9.80	1.25	0.08	0.07	0.11	0.36	0.09	0.06	0.05	0.08	1.6
<b>F58L</b>	779	4.50	2.19	0.01	0.02	0.02	0.04	0.02	0.01	0.01	0.01	1.54
<b>F94L</b>	581	13.00	1.42	0.06	0.17	0.13	0.27	0.12	0.08	0.12	0.1	1.64
<b>G24D</b>	840	12.20	1.27	0.02	0.04	0.12	0.14	0.08	0.03	0.05	0.04	1.56
<b>G78C</b>	601	12.60	1.17	0.06	0.03	0.07	0.11	0.06	0.06	0.07	0.05	1.58
<b>H51R</b>	953	15.00	1.51	0.06	0.06	0.07	0.27	0.09	0.07	0.1	0.07	1.65
<b>H57R</b>	501	10.40	1.61	0.04	0.03	0.03	0.14	0.10	0.04	0.06	0.05	1.54
<b>H71Y</b>	636	12.70	1.99	0.02	0.02	0.01	0.11	0.04	0.03	0.02	0.02	1.53
<b>K48T</b>	466	19.60	1.42	0.02	0.01	0.01	0.02	0.02	0.03	0.02	0.02	1.61
<b>L116P</b>	1265	11.20	1.62	0.03	0.05	0.05	0.1	0.06	0.03	0.05	0.05	1.58
<b>L4S</b>	1536	16.30	1.54	0.04	0.07	0.08	0.2	0.07	0.04	0.04	0.04	1.63
<b>P54L</b>	0	10.10	1.95	0.01	0.03	0.06	0.09	0.11	0.03	0.03	0.02	1.53
<b>P62L</b>	0	9.20	1.84	0.01	0.02	0.01	0.06	0.01	0.02	0.1	0.03	1.54
<b>Q10P</b>	587	9.80	1.84	0.01	0.04	0.08	0.22	0.02	0.01	0.04	0.04	1.53
<b>R29P</b>	718	13.50	1.21	0.01	0.03	0.11	0.17	0.05	0.01	0.02	0.02	1.56
<b>T135P</b>	744	15.50	2.12	0.01	0.01	0.03	0.18	0.05	0.02	0.03	0.03	1.53
<b>T142A</b>	623	10.10	1.63	0.01	0.02	0.03	0.24	0.01	0.01	0.01	0.01	1.54
<b>T160K</b>	412	5.90	2.12	0.03	0.02	0.02	0.07	0.02	0.03	0.03	0.02	1.52
<b>T76P</b>	851	9.6	2.02	0.02	0.04	0.02	0.19	0.04	0.03	0.05	0.04	1.58
<b>V139A</b>	597	10.6	1.49	0.02	0.04	0.07	0.18	0.05	0.03	0.04	0.03	1.54
<b>W119L</b>	1068	14.2	2.07	0.02	0.02	0.02	0.07	0.01	0.04	0.04	0.03	1.58
<b>Y34D</b>	657	4.6	1.7	0.03	0.01	0.03	0.03	0.02	0.03	0.03	0.02	1.59
<b>Y64D</b>	1575	17.4	1.97	0.06	0.08	0.09	0.25	0.16	0.07	0.07	0.06	1.7
<b>A171T</b>	521	13.1	2.07	0.01	0.03	0.03	0.08	0.05	0.02	0.03	0.01	1.53
<b>A46V</b>	507	17.7	1.51	0.09	0.1	0.09	0.48	0.13	0.06	0.08	0.06	1.62
<b>L172P</b>	695	13.3	1.76	0.02	0.05	0.04	0.12	0.16	0.02	0.06	0.04	1.52
<b>M175V</b>	1579	16.4	1.29	0.05	0.07	0.22	0.19	0.05	0.04	0.05	0.05	1.63
<b>P62R</b>	1019	14.4	1.58	0.02	0.04	0.07	0.23	0.08	0.06	0.12	0.04	1.57
<b>V125F</b>	969	18.4	1.33	0.09	0.11	0.14	0.36	0.32	0.08	0.17	0.08	1.62
<b>V180F</b>	835	11.7	1.49	0.01	0.02	0.03	0.03	0.04	0.02	0.02	0.01	1.57
<b>Desv. St.</b>	371.9	3.91	0.41	0.03	0.03	0.05	0.11	0.07	0.02	0.04	0.02	0.04



**Figura 19.** Fluctuación del volumen del sitio catalítico del modelo de la enzima *PZasa wild type* y de las mutaciones más representativas.

En contraste, los parámetros fisicoquímicos determinados fueron los que presentaron menores fluctuaciones. De entre estos, la mayor varianza ha sido determinado para el SASA. Particularmente, L116P, V180F y M175V presentaron los mayores valores de SASA. Además, la mutación Q10P ocasiona la disminución del SASA, siendo inferior al valor encontrado en el *wild type*. En cuanto al PI, II y AI, los valores presentaron fluctuaciones mínimas (Tabla 5).

**Tabla 5.** Parámetros fisicoquímicos de la *PZasa wild type* y mutada de *M. tuberculosis*. Los datos son obtenidos de las estructuras analizadas en el PCA.

Mutación	SASA	PI	II	AI
WT	9655.13	4.43	9.32	84.38
A102V	8591.46	4.43	9.37	85.41
C14G	9234.60	4.43	8.91	84.38
D12A	8991.07	4.48	9.72	84.92

<b>D12G</b>	9102.81	4.48	8.91	84.38
<b>D136G</b>	9489.36	4.48	8.86	84.38
<b>D49N</b>	9020.83	4.48	8.91	84.38
<b>F58L</b>	9095.29	4.43	9.88	86.49
<b>F94L</b>	9506.40	4.43	7.55	86.49
<b>G24D</b>	9034.72	4.39	9.11	84.38
<b>G78C</b>	9225.74	4.43	9.37	84.38
<b>H51R</b>	9355.25	4.43	6.94	84.38
<b>H57R</b>	8941.72	4.43	9.47	84.38
<b>H71Y</b>	8824.74	4.37	9.32	84.38
<b>K48T</b>	9036.67	4.36	9.32	84.38
<b>L116P</b>	9258.19	4.43	10.36	82.27
<b>L4S</b>	9173.56	4.43	9.32	82.27
<b>P54L</b>	9086.40	4.43	9.32	86.49
<b>P62L</b>	8851.84	4.43	9.72	86.49
<b>Q10P</b>	8697.56	4.43	10.36	84.38
<b>R29P</b>	9004.27	4.36	11.4	84.38
<b>T135P</b>	8992.19	4.43	9.95	84.38
<b>T142A</b>	8775.36	4.43	9.32	84.92
<b>T160K</b>	8680.51	4.5	8.86	84.38
<b>T76P</b>	9005.90	4.43	10.82	84.38
<b>V139A</b>	9017.13	4.43	9.72	83.35
<b>W119L</b>	9132.18	4.43	9.21	86.49
<b>Y34D</b>	9118.79	4.39	9.32	84.38
<b>Y64D</b>	9163.87	4.39	10.36	84.38
<b>A171T</b>	9008.96	4.43	9.32	83.84
<b>A46V</b>	9098.59	4.43	8.86	85.41
<b>L172P</b>	8902.96	4.43	11.3	82.27
<b>M175V</b>	9290.50	4.43	9.72	85.95
<b>P62R</b>	9401.46	4.5	9.72	84.38
<b>V125F</b>	9113.96	4.43	10.8	82.81
<b>V180F</b>	9203.68	4.43	9.32	82.81
<b>Desv. St.</b>	230.85	0.03	0.87	1.15

#### 4.8 Predicción de la actividad enzimática

El análisis de regresión lineal múltiple sugiere que cuatro parámetros son los mejores predictores para la actividad enzimática. Entre estos, el radio de giro ( $R_{gx}$ ), RMSF del residuo 138 (uno de los principales residuos que participa en la actividad enzimática), RMSF del residuo 57 (el cual uno de los cuatro residuos que coordina con el cofactor metálico). Cabe señalar que todos ellos presentan un *p-value* significativo, sin embargo, el valor del  $r^2 = 0.48$  (I.C. 95% = 0.26 - 0.77) sugiere que el modelo no resultaría en un buen predictor de la actividad enzimática (Tabla 6). Por otro lado, los modelos de  $Eff$ ,  $K_{cat}$  y  $K_m$ , presentaron valores de  $r^2$  fueron más bajos ( $r^2 > 0.19$ ) para poder ser usado como un modelo de predicción.

**Tabla 6.** Análisis de regresión lineal múltiple entre los parámetros geométricos y fisicoquímicos para predecir la actividad enzimática de la *PZasa* de *M. tuberculosis*.

Model	Variable	Coeficiente	Std. Error	F-test	<i>p-value</i>	I.C. (95%)	
						2.5%	97.5%
	II	23.388	7.953	8.6482	0.007	6.973	39.802
	RGX	400.849	169.875	5.5680	0.027	50.243	751.454
Act.	RMSF-138	11156.166	471.770	6.0059	0.021	182.480	2129.852
	RMSF-57	-252.103	90.251	7.8027	0.010	-438.372	-65.832
	Constant	-828.391	270.577	9.3732	0.0053	-1386.834	-269.94649
<b>Coefficiente de correlación (<math>r^2</math>):</b>					<b>0.48*</b>	<b>0.26</b>	<b>0.77</b>

\*El I.C. 95% para el valor de  $r^2$  fue estimado mediante re-muestreo con 5000 muestras.

Por otra parte, en los modelos usando transformaciones logarítmicas encontramos mayores correlaciones en comparación al primer modelo de regresión lineal múltiple. En este análisis estimamos mayores valores de  $r^2$  para tres parámetros

cinéticos (Tabla 7). El mejor modelo se obtuvo para predecir el  $K_{cat}$ , con un  $r^2 = 0.68$  (I.C. 95% = 0.53 - 0.89). Para la  $Act.$  y el  $K_m$  los modelos también presentaron altos coeficientes de correlación,  $r^2 = 0.67$  (I.C. 95% = 0.48 - 0.89) y  $r^2 = 0.65$  (I.C. 95% = 0.50 - 0.90), respectivamente. El modelo con menor correlación fue el de  $Eff.$ , con un  $r^2 = 0.29$  (I.C. 95% = 0.04 - 0.61). Los tres modelos mencionados inicialmente superan a los modelos que no han sido trabajados con datos transformados, validando que la metodología aplicada resultó útil para la construcción de los modelos predictivos.

Asimismo, se identificó una marcada diferencia en los residuos que contribuyen a describir estos nuevos modelos en comparación a los seleccionados para el análisis anterior. Entre ellos, se observó que el RMSF de ciertos residuos contribuyen significativamente con el modelo. Cabe resaltar que para los modelos con las transformaciones logarítmicas no se observó relación entre los residuos que participan directamente con la actividad enzimática, con excepción del modelo del  $K_{cat}$ . en el que el RMSF del residuo 138 resulto ser significativo (Tabla 7).

**Tabla 7.** Modelos para los parámetros cinéticos transformados y los parámetros de estabilidad, geométricos y fisicoquímicos.

Modelo_log10	Variable	Coeficiente	Std. Error	F-test	p-value	I.C. (95%)	
						2.5%	97.5%
<i>Act._log</i>	Vol.	-0.001	0.000	6.315	0.019	-0.002	-0.0002
	II	1.075	0.218	24.431	0.0001	0.625	1.525
	RMSF8	30.133	8.821	11.670	0.0024	11.886	48.378
	RMSF49	27.801	6.561	17.957	0.0003	14.227	41.370
	RMSF57	-12.609	2.347	28.854	0.00001	-17.463	-7.752
	Constant	-8.821	2.101	17.629	0.0003	-13.167	-4.475
<b>Coefficiente de correlación (r<sup>2</sup>)</b>					<b>0.67</b>	<b>0.48</b>	<b>0.89</b>
<i>Kcat._log</i>	II	0.476	0.124	14.649	0.0009	0.217	0.731
	RMSF138	23.920	7.161	11.157	0.0028	9.060	38.647
	RMSF57	-6.390	1.413	20.462	0.0002	-9.305	-3.468
	RMSF71	7.083	2.165	10.705	0.0034	2.623	11.567
	Vol.	-0.001	0.000	8.790	0.0069	-0.001	-0.0002
	Constant	-4.161	1.246	11.157	0.0028	-6.721	-1.575
<b>Coefficiente de correlación (r<sup>2</sup>)</b>					<b>0.68*</b>	<b>0.53</b>	<b>0.89</b>
<i>Eff._log</i>	AI	0.5086	0.2421	2.100	0.0459	0.009	1.007
	II	0.8472	0.3038	2.789	0.0099	0.221	1.472
	Constant	-50.3329	21.2249	-2.371	0.0257	-94.046	-6.619
<b>Coefficiente de correlación (r<sup>2</sup>)</b>					<b>0.29*</b>	<b>0.04</b>	<b>0.61</b>
<i>Km_log</i>	RMSF8	11.362	2.721	17.437	0.0004	5.718	17.001
	RMSF51	-7.628	1.924	15.712	0.0007	-1.161	-3.637
	RMSF124	30.395	6.865	19.601	0.0002	1.614	44.614
	RMSF173	-12.931	3.192	16.409	0.0005	-1.954	-6.303
	RMSF185	3.599	0.893	16.254	0.0006	1.748	5.450
	Weight	0.004	0.002	6.719	0.0166	8.440	0.007
	Constant	-82.466	31.925	6.673	0.0170	-1.486	-16.245
<b>Coefficiente de correlación (r<sup>2</sup>)</b>					<b>0.65*</b>	<b>0.50</b>	<b>0.90</b>

\*Los I.C. 95% para los valores de r<sup>2</sup> fueron estimados mediante re-muestreo con 5000 muestras.

#### 4.9 Especificidad y sensibilidad

La estimación de la especificidad y sensibilidad se realizó con el modelo construido para la actividad enzimática, debido a que es el más importante en la determinación de cepas resistentes (Tabla 8 y 6). Para la predicción de las cepas sensibles se usó un *cut-off* de 90%, mientras que para las cepas resistentes se consideraron aquellas con actividad <90.0%. Los valores verdaderos para las cepas resistentes (VP) fueron obtenidos con las mutaciones A171T, A46V, M175V, P62R y V180F; mientras que, los falsos positivos fueron observados en las mutaciones L172P y V125F. Por otro lado, determinamos que 3 verdaderos negativos (VN) con el PC1, PC2 y PC6, que corresponden a estructuras *wild type* y 4 valores falsos negativos (FN) en PC3, PC4, PC5 y PC7 (Tabla 8). Asimismo, los análisis de la matriz de confusión muestran que la sensibilidad (55.56% / IC. 95% 26.66-81.12) y especificidad (60.0% / IC. 95% 23.07-88.24) tendrían significancia media para clasificar correctamente los casos de TB resistente (Tabla 9).

**Tabla 8.** Predicción de la actividad enzimática con data no ingresada en el modelo.

Mutac.	Act. (%)	Resist .	Vol	II	RMSF			Act. Predicha*	Clase
					8	49	57		
A171T	55.18	SI	521	9.32	0.01	0.03	0.08	0.80364	VP
A46V	41.41	SI	507	8.86	0.09	0.1	0.48	-0.36375	VP
L172P	1.81	SI	695	11.3	0.02	0.05	0.12	3.11113	FP
M175V	20.92	SI	1579	9.72	0.05	0.07	0.19	1.10601	VP
P62R	28.57	SI	1019	9.72	0.02	0.04	0.23	-0.57637	VP
V125F	72.43	SI	969	10.8	0.09	0.11	0.36	3.05084	FP
V180F	32.98	SI	835	9.32	0.01	0.02	0.03	0.84208	VP
WT-PC1	100.00	NO	553	9.32	0.386	0.0616	0.17	11.8453	VN
WT-PC2	100.00	NO	583	9.32	0.0153	0.0680	0.39	-1.95100	VN
WT-PC3	100.00	NO	521	9.32	0.0358	0.0548	0.36	-1.25998	FN

WT-PC4	100.00	NO	635	9.32	0.102	0.0551	0.34	0.88134	FN
WT-PC5	100.00	NO	618	9.32	0.102	0.0648	0.31	1.54628	FN
WT-PC6	100.00	NO	601	9.32	0.210	0.0625	0.29	5.00588	VN
WT-PC7	100.00	NO	587	9.32	0.031	0.0557	0.31	-0.8151	FN

VP= verdaderos positivos; FP= falsos positivos; VN= verdaderos negativos; FN= falsos negativos.  
 \*Debido a que el modelo se ha trabajado con datos transformados, se ha utilizado como cut-off el nivel de 90% de actividad para que sea resistente, en tal sentido su valor correspondiente de 1.95 es el cut-off que determina si es sensible o resistente.

**Tabla 9.** Matriz de confusión.

		PREDICCIÓN	
		POSITIVOS (RESISTENTES)	NEGATIVOS (SENSIBLES)
OBSERVACIÓN	POSITIVOS	5 (VP)	2 (FP)
	NEGATIVOS	4 (FN)	3 (VN)
SENSIBILIDAD		55.56%	IC. 95% (26.66-81.12)
ESPECIFICIDAD		60.00%	IC. 95% (23.07-88.24)

## V. DISCUSIÓN

Mutaciones puntuales en la enzima pirazinamidasa han sido señaladas como causantes de resistencia a pirazinamida, en *M. tuberculosis*. Se reporta que el 70% de MDR y 96% de XDR-TB están asociadas con mutaciones en este gen *pncA* (56). En este estudio se han tomado parámetros enzimáticos (*Act. Kcat.*, *Km.*, *Eff.*) que describen el *fitness* de 35 cepas mutantes y un *wild type* (24). Sin embargo, hasta el momento las modificaciones a nivel de conformación no han podido ser determinadas a detalle, debido a las limitaciones de resolución a escala atómica. Es por ello que, en esta investigación se ha logrado describir a nivel atómico la influencia de cada mutación sobre la estructura y la dinámica molecular de la enzima pirazinamidasa.

En este estudio, clasificamos a las mutaciones de acuerdo a su localización en la conformación tridimensional de la enzima. Las mutaciones que estuvieron cercanas al sitio de coordinación metálico están comprendidas entre los residuos Asp49 a Ser74, en donde 4 de estos residuos (Asp49, His51, His57 e His71) conforman una tetraedra de coordinación con su cofactor metálico, el cual puede ser Fe o Zn (57).

Por otro lado, las mutaciones que se localizan dentro del sitio catalítico comprenden una cavidad conformada por 19 residuos, en donde de Asp8, Ile133 y Cys138 participan en la conversión de pirazinamida a ácido pirazinóico (41). Cabe resaltar que ninguna de las mutaciones reportadas modifica a los residuos catalíticos, lo cual es consistente con la pérdida de el *fitness* y no de la totalidad de la actividad enzimática. Asimismo, mediante esta clasificación ha sido posible observar

patrones de modificaciones en las diferentes estructurales, en las simulaciones de dinámica molecular principalmente.

En relación a los modelos mutados, a nivel estructural estos resultaron altamente similares entre sí, determinándose valores menores a 0.0001 en el RMSD. Estas leves fluctuaciones son atribuidas al tipo de algoritmo usado para mutar y refinar las estructuras, el cual solo realiza un cambio de aminoácido seguido de una optimización mediante una dinámica molecular rápida (58). En ese sentido, mediante esta metodología no es posible determinar cambios sustanciales en la conformación de la proteína, lo cual resulta una limitación para que los cálculos de acoplamiento molecular determinen correctamente el modo y energía de unión del complejo fármaco-proteína. Otros métodos que exploran los efectos de las mutaciones han sido desarrollados, como por ejemplo NMA (59), sin embargo, estos presentan la limitación de la exploración conformación, siendo limitando el espacio de búsqueda. Es por ello que consideramos más apropiado realizar simulaciones de dinámica molecular para tal fin.

Dentro de los acoplamientos moleculares, determinamos que las ligeras fluctuaciones en los valores de *binding energy* ( $\text{Kcal mol}^{-1}$ ) no resultan significantes para ser asociados con las fluctuaciones observadas en los parámetros cinéticos experimentales. Sin embargo, se ha observado que en la estructura mutada H57R se presentó un modo de interacción particular entre la droga y el residuo mutado (Arg 57), al mismo tiempo se observó la pérdida de interacciones con residuos catalíticos importantes. Ello puede relacionarse con el descenso de la actividad enzimática por la pérdida de afinidad con el cofactor en este sitio y la interacción con la droga.

Asimismo, resulta importante señalar que las diferentes metodologías de docking molecular empleadas presentan limitaciones con relación a la solubilidad del ligando, en tal sentido nuestros análisis sugieren que los mejores modelos son los obtenidos con el método hidratado, con el cual se ha podido observar que moléculas de agua contribuyen a la estabilidad de la droga. Estas interacciones son similares a lo que se ha reportado en la resolución cristalográfica de la enzima. Futuros trabajos pueden beneficiarse de estas metodologías para explorar con mayor precisión el acoplamiento con otras mutaciones.

Es sabido que, las mutaciones puntuales en las secuencias proteicas pueden modificar la estabilidad y la función de las proteínas (60); en tal sentido, determinar sus efectos ha sido uno de los principales retos en el campo de la biología estructural. Diversos estudios señalan que la simulación de dinámica molecular como método óptimo para determinar el efecto de las mutaciones (52,61,62). Además, mediante simulaciones de dinámica molecular se pueden terminar los efectos de mutaciones a corto y largo alcance.

Cabe resaltar que los análisis de las dinámicas moleculares sugieren un amplio cambio conformacional entre las diferentes estructuras mutadas. Ello resulta interesante debido a que tal modificación en la estructura es causada por una sola mutación puntual. Como se ha observado, el plot de RMSD indica que la estructura *wild type* converge a partir de los 100 ns, estimándose un tiempo de simulación adecuado. No obstante, las estructuras mutadas presentan diversos patrones en los valores de RMSD, y en alguna de ellas el tiempo de simulación puede no ser el adecuado para determinar una convergencia total del sistema, como en la estructura mutada Y64D en donde no se observa una convergencia total. Esto puede incidir en

los cálculos de los parámetros geométricos y fisicoquímicos, como se ha reportado previamente (63).

Nuestros tiempos de simulación, aunque son relativamente extensos (500 ns), pueden resultar cortos para determinar modificaciones estructurales importantes como resultados de las mutaciones. Esto limita la precisión en los casos en los que los sistemas no convergen, sin embargo, cabe la posibilidad que estos no alcancen dicho estado como un fenómeno propio de las mutaciones. En ese sentido, extender las simulaciones de dinámica molecular a escalas de tiempo de microsegundos ( $\mu\text{s}$ ) o milisegundos (ms) podría resultar beneficiosos para mejorar nuestros análisis, a pesar de que esto representa un gran reto computacional. Una metodología de *Accelerated Molecular Dynamics* (aMD) (64) podría reducir el coste computacional mejorando el muestreo del espacio conformacional mediante la reducción de las barreras energéticas que separan a cada estado.

Por otro lado, los valores de RMSF fueron más exactos para determinar fluctuaciones en los aminoácidos mutados y otros importantes. Interesantemente, la estructura mutada H57R presentó una marcada disminución en el RMSF, ello es consistente con los resultados obtenidos en el docking molecular, asimismo, se sugiere que esta mutación estaría causando una menor actividad catalítica por un desfavorecimiento de unión del cofactor metálico, lo cual ha sido reportado en otro trabajo como un factor que reduce la actividad enzimática (57).

Además, es importante señalar que la región del loop que conforma este residuo fue la más variable. Por su parte, las mutaciones K48T y Y54D que tienen altos valores de RMSF ocasiona una apertura del sitio catalítico que es consistente con los

mayores volúmenes observados en estas estructuras, evidenciándose una relación entre estos dos parámetros. Por lo tanto, estas mutaciones podrían ser consideradas como las mutaciones más importantes en esta proteína y para futuros análisis de resistencia a pirazinamida.

En cuanto al análisis de componente principales, las mayores fluctuaciones observadas en cada mutación pueden estar relacionadas con modificaciones alostéricas que se dan a largo alcance. Una modificación alostérica puede perturbar la conformación de la estructura a distancias superiores de los 10 Å (65), tal y como se ha evidenciado en nuestros análisis. Estas modificaciones a largo alcance se resaltan en las mutaciones H51R, Y64D, K48T, D136G, Y34D y M175V, permitiendo estimar parámetros estructurales marcadamente diferenciados, lo cuales se asocian con aperturas y cerraduras del sitio catalítico.

En otros trabajos se ha asociado la pérdida de la actividad de fármacos debido a este factor, el cual ocasiona, entre otras cosas, incremento del volumen de la cavidad del sitio activo (66,67). Es importante resaltar que, aunque existen mutaciones que ocurren fuera del sitio catalítico y de unión al cofactor metálico, se evidencia un efecto marcado sobre estos sitios. Ello resulta importante para evaluar y asociar mutaciones en todas las regiones de la proteína sobre la conformación tridimensional de la estructura.

En las observaciones de las simulaciones de dinámica molecular se ha podido identificar que las mutaciones ocasionan 2 eventos importantes, la apertura y el cierre del sitio catalítico. Estos eventos resultan en modificaciones de la geometría natural de la enzima *PZasa*. Considerando ello, podemos asociar algunas cepas

resistentes al cierre, como A102V, F58L, H57R, H71R, P54L, P62L, T76P, C14G, D12A, D12G, G78C, L116P, R29P, T135P, T160K, V139A, Y34D, A171T y L172P; mientras que H51R, Y64D, P62R, D136G, D49N, K48T, Q10P, F94L, G24D, T142A, L4S, W119L, A46V, M175V, V125F y V180F ocasionan aperturas en el sitio catalítico que desfavorecen la interacción con la droga pirazinamida.

Los parámetros estructurales fueron más fluctuantes que los observados en los parámetros fisicoquímicos. Esto se puede atribuir a que el método de PCA trata de identificar principalmente cambios a nivel de conformación; además, los métodos computacionales a nivel de dinámica molecular clásica son limitados para estimar reacciones químicas, por lo que resulta necesario evaluar otras metodologías que puedan modelar la reacción química que se daría con las diferentes estructuras mutadas y la droga pirazinamida en el sitio activo. Para ello, se puede hacer uso de métodos cuánticos (QM) y/o métodos híbridos de dinámica molecular (QM/MM) (68), mediante las cuales puede estimarse la formación y/o rompimiento de enlaces químicos, y para nuestro enfoque determinar la conversión de pirazinamida en ácido pirazinóico. Sin embargo, estos presentan la desventaja de requerir gran capacidad de cómputo.

En cuanto a los modelos de predicción con regresión lineal múltiple, solo se logró construir un modelo significativo para la Actividad enzimática (*Act. %*), el cual presentó un valor de  $r^2 = 0.48$ . Sin embargo, en los modelos con los valores transformados, se ha observado una mejor predicción en todos los parámetros cinéticos, mediante la toma de la estabilidad de las estructuras mutadas durante toda la dinámica. Ello sugiere que la selección de un mayor número de *frames* mejora los análisis predictivos, en comparación del método de PCA que reduce la cantidad

de *frames*. Además, la diferencia entre los residuos importantes para la construcción del modelo sugiere que las mutaciones tienen una marcada influencia sobre residuos distantes al sitio activo y de coordinación metálico. Ello también se relaciona con nuestros resultados obtenidos en el PCA, en donde observamos cambios en regiones distantes a ambos sitios.

Cabe señalar que una investigación reciente usando modos normales de vibración (NMA) se obtuvo un modelo de predicción con un valor de  $r^2$  de 0.897 para el  $K_{cat}$  (69). En dicho estudio se encontró un RMSD de 0.094 Å en las estructuras mutadas, el cual es superior a nuestros modelos mutados puntualmente. Además, estimando los valores del RMSF, se identificó el mayor pico de fluctuación en los residuos His49 e His71; sin embargo, se observó que dichas fluctuaciones son menores en los modelos obtenidos por modos normales de vibración (NMA) en comparación a los obtenidos por dinámica molecular (MD).

Estos resultados nos llevan a pensar que usando los datos completos de la dinámica molecular se pueden identificar mayores cambios estructurales que están asociados con la función de la enzimática. Sin embargo, los modos normales son más precisos para encontrar una relación entre la estructura y la función enzimática, especialmente el  $K_{cat}$ . Más simulaciones son requeridas para validar esta teoría.

Cabe señalar que el método de NMA consideró 185 descriptores fisicoquímicos con 40 modelos de vibración generados, mientras que nuestro análisis de PCA generó 30 modelos, pero en los cuales se identificó mayor fluctuación tanto de RMSD y RMSF. Asimismo, en las estimaciones de estabilidad con el método de NMA en las mutaciones que presentan  $K_{cat}$  mayor o igual al 50%, no se observan diferencias

significativas. Por su parte en aquellas que tienen un valor menor al 50% se observan diferencias significativas en las posiciones Leu27, Pro70, Tyr103, Val109 y Ser179. Comparando a nuestros resultados, los análisis de PCA detectan fluctuaciones marcadamente diferentes en todas las mutaciones, habiendo mayores modificaciones conformacionales en las mutaciones C14G, D12A, D12G y Q10P, que pertenecen al sitio catalítico.

El menor valor de  $r^2$  en el primer modelo puede atribuirse a la poca cantidad de datos estimados (12 parámetros geométricos y 4 fisicoquímicos), dejando abierta la posibilidad de extender las simulaciones y estimar más parámetros estructurales para cada uno de los *frames*, con lo cual se podría conseguir un mejor modelo predictivo. Por su parte los parámetros de *Kcat*, *Km* y *Eff* son muy bajos para establecer una correlación lineal. La aplicación de un modelo no lineal u otros basados en inteligencia artificial podrían ser considerados para estimar la relación entre estos parámetros con la finalidad de mejorar la predicción.

Por otro lado, los modelos trabajados con datos transformados logarítmicamente muestran un mejor coeficiente de correlación para predecir la *Act.*, *Kcat* y *Eff*. Realizar transformaciones logarítmicas sobre bases de datos biológicos resulta interesante puesto que muchos de los eventos biológicos se suelen presentar en escala logarítmica. Además, se sabe que realizar transformaciones logarítmicas puede ser conveniente para normalizar la data de los experimentos (69). Esto explicaría los buenos coeficientes de correlación obtenidos con este método.

La sensibilidad que refleja la habilidad del modelo logarítmico para detectar los casos relevantes y la especificidad que representa la proporción entre los casos

negativos clasificados correctamente, fueron estimados como aceptables. Sin embargo, estos valores son bajos para garantizar una correcta discriminación entre las cepas resistentes. Es importante señalar que no se han encontrado modelos de predicción enzimática que evalúen tales parámetros. Por lo que nuestro flujo de trabajo puede ser considerado como base y mejorado en futuras investigaciones que tomen más data para mejorar las predicciones usando métodos de simulación de biomoléculas.

Finalmente, la poca cantidad de parámetros cinéticos para las cepas resistentes a pirazinamida resulta una limitante para poder realizar análisis de sensibilidad y especificidad más precisos. Se conocen por lo menos 641 mutaciones en 171 codones de gen *pncA* que causan resistencia a pirazinamida (70), sin embargo, en la mayoría de trabajos disponibles solo se han realizado ensayos de resistencia en sembrado en placas, quedando la necesidad de estimar los parámetros cinéticos para todas estas variantes. De estimarse estos valores experimentalmente, dicha información podría ser ingresada a nuestros modelos para mejorar la precisión de los cálculos.

## VI. CONCLUSIONES

- Las metodologías de docking molecular, con y sin ion metálico (Fe), hidratado y no hidratado, no difieren significativamente en el cálculo de las energías de unión ( $\text{Kcal mol}^{-1}$ ) con respecto al modelo *wild type*. Sin embargo, se ha podido identificar un modo de acoplamiento diferente en la mutación H57R, el cual destaca por ser único entre todas las estructuras mutadas.
- Los análisis de dinámica molecular y el análisis de los componentes principales de la enzima *PZasa* corroboran el efecto que ejerce una única mutación puntual, a nivel de conformación de la estructura. En dichos análisis los valores de RMSD y RMSF fluctúan significativamente. Asimismo, la variabilidad en los parámetros geométricos y fisicoquímicos observados indican mayores y menores efectos que son causados por cada mutación en particular.
- Los mejores modelos de predicción fueron obtenidos con los datos experimentales transformados logarítmicamente. El modelo con mayor coeficiente de correlación fue el que describe el *Kcat* ( $r^2 = 0.68$  / I.C. 95% = 0.53 - 0.89). Los modelos de *Km* y *Act.* también tuvieron altos coeficientes de correlación ( $r^2 = 0.67$  / I.C. 95% = 0.48 - 0.89 y  $r^2 = 0.65$  / I.C. 95% = 0.50 - 0.90, respectivamente). El más bajo coeficiente se obtuvo con la *Eff.* ( $r^2 = 0.29$  / I.C. 95% = 0.04 - 0.61).
- Estos modelos también sugieren que los parámetros de estabilidad (RMSF) de los residuos H51 y H71 son importantes por su significancia en estos modelos. Los términos seleccionados en estos modelos están relacionados a la

interacción con el cofactor metálico, por lo que resulta ser un evento importante en la resistencia de la *PZasa*.

## VII. RECOMENDACIONES

- Implementar otras metodologías de docking molecular en donde se aplica libertad a las torsiones del sitio catalítico con la finalidad de ampliar el espacio conformacional y la búsqueda de una pose correcta.
- Ampliar los tiempos de simulación con la finalidad de determinar mayores ensambles de conformómeros que puedan ser usados para determinar mayores parámetros geométricos y fisicoquímicos.
- Realizar una simulación de dinámica molecular tomando al ion metálico y a la droga para cada una de los sistemas construidos con las mutaciones descritas.
- Incluir otras mutaciones que dispongan de data experimental para mejorar la predicción del modelo.
- Evaluar otros tipos de funciones no lineales que puedan explicar mejor el modelo predictivo de la actividad enzimática.

## VIII. REFERENCIAS

1. Pięłowski M. Pathogenic and Non-Pathogenic Microorganisms in the Rapid Alert System for Food and Feed. *Int J Environ Res Public Health* [Internet]. febrero de 2019 [citado 4 de febrero de 2021];16(3). Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6388125/>
2. Waldman AJ, Balskus EP. The Human Microbiota, Infectious Disease, and Global Health: Challenges and Opportunities. *ACS Infect Dis*. 12 de enero de 2018;4(1):14-26.
3. Ramazanzadeh R, Roshani D, Shakib P, Rouhi S. Prevalence and occurrence rate of Mycobacterium tuberculosis Haarlem family multi-drug resistant in the worldwide population: A systematic review and meta-analysis. *J Res Med Sci* [Internet]. enero de 2015 [citado 4 de febrero de 2021];20(1):78-88. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4354070/>
4. Organization WH. Global tuberculosis report 2018 [Internet]. World Health Organization; 2018 [citado 6 de febrero de 2021]. Disponible en: <https://apps.who.int/iris/handle/10665/274453>
5. Chauhan V, Thakur S. State of the Globe: The Global Battle for Survival against Mycobacterium tuberculosis. *J Glob Infect Dis* [Internet]. 2017 [citado 4 de febrero de 2021];9(4):129-30. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5750434/>
6. Calixto RP. OPS/OMS Perú [Internet]. Pan American Health Organization / World Health Organization. 2018 [citado 4 de febrero de 2021]. Disponible en: [https://www.paho.org/per/index.php?option=com\\_content&view=article&id=4075:tuberculosis&Itemid=0](https://www.paho.org/per/index.php?option=com_content&view=article&id=4075:tuberculosis&Itemid=0)
7. Soto Cabezas MG, Munayco Escate CV, Escalante Maldonado O, Valencia Torres E, Arica Gutiérrez J, Yagui Moscoso MJA. Perfil epidemiológico de la tuberculosis extensivamente resistente en el Perú, 2013-2015. *Rev Panam Salud Publica* [Internet]. 23 de septiembre de 2020 [citado 4 de febrero de 2021];44. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7498293/>
8. Brossier F, Boudinet M, Jarlier V, Petrella S, Sougakoff W. Comparative study of enzymatic activities of new KatG mutants from low- and high-level isoniazid-resistant clinical isolates of Mycobacterium tuberculosis. *Tuberculosis (Edinb)*. septiembre de 2016;100:15-24.
9. Saukkonen J. Rifampin and Pyrazinamide for Latent Tuberculosis Infection: Clinical Trials and General Practice. *Clinical Infectious Diseases* [Internet]. 15 de agosto de 2004 [citado 4 de febrero de 2021];39(4):566-8. Disponible en: <https://doi.org/10.1086/422728>

10. Zhang Y, Mitchison D. The curious characteristics of pyrazinamide: a review. *Int J Tuberc Lung Dis.* enero de 2003;7(1):6-21.
11. McGrath M, Gey van Pittius NC, van Helden PD, Warren RM, Warner DF. Mutation rate and the emergence of drug resistance in *Mycobacterium tuberculosis*. *J Antimicrob Chemother.* febrero de 2014;69(2):292-302.
12. Zhang Y, Yew W-W. Mechanisms of drug resistance in *Mycobacterium tuberculosis*: update 2015. *Int J Tuberc Lung Dis.* noviembre de 2015;19(11):1276-89.
13. Medina E, Pieper DH. Tackling Threats and Future Problems of Multidrug-Resistant Bacteria. *Curr Top Microbiol Immunol.* 2016;398:3-33.
14. Ferri M, Ranucci E, Romagnoli P, Giaccone V. Antimicrobial resistance: A global emerging threat to public health systems. *Critical Reviews in Food Science and Nutrition* [Internet]. 2 de septiembre de 2017 [citado 1 de septiembre de 2021];57(13):2857-76. Disponible en: <https://www.tandfonline.com/doi/full/10.1080/10408398.2015.1077192>
15. Protein dynamics and motions in relation to their functions: several case studies and the underlying mechanisms - PubMed [Internet]. [citado 6 de febrero de 2021]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/23527883/>
16. Leonard PJ, Rathod PK, Golin J. Loss of function mutation in the yeast multiple drug resistance gene PDR5 causes a reduction in chloramphenicol efflux. *Antimicrob Agents Chemother.* octubre de 1994;38(10):2492-4.
17. Agarwal PK. Enzymes: An integrated view of structure, dynamics and function. *Microb Cell Fact* [Internet]. diciembre de 2006 [citado 1 de septiembre de 2021];5(1):2. Disponible en: <https://microbialcellfactories.biomedcentral.com/articles/10.1186/1475-2859-5-2>
18. Banerjee P, Siramshetty VB, Drwal MN, Preissner R. Computational methods for prediction of in vitro effects of new chemical structures. *J Cheminform* [Internet]. diciembre de 2016 [citado 1 de septiembre de 2021];8(1):51. Disponible en: <https://jcheminf.biomedcentral.com/articles/10.1186/s13321-016-0162-2>
19. Kuznetsov V, Lee HK, Maurer-Stroh S, Molnár MJ, Pongor S, Eisenhaber B, et al. How bioinformatics influences health informatics: usage of biomolecular sequences, expression profiles and automated microscopic image analyses for clinical needs and public health. *Health Inf Sci Syst* [Internet]. 10 de enero de 2013 [citado 6 de febrero de 2021];1. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4336111/>
20. Muhammad AM, Ujala S, Niaz A. Use of Bioinformatics Tools in Different Spheres of Life Sciences. *Journal of Data Mining in Genomics & Proteomics*

- [Internet]. 2014 [citado 6 de febrero de 2021]; Disponible en: <https://www.readcube.com/articles/10.4172%2F2153-0602.1000158>
21. Borhani DW, Shaw DE. The future of molecular dynamics simulations in drug discovery. *J Comput Aided Mol Des* [Internet]. enero de 2012 [citado 6 de febrero de 2021];26(1):15-26. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3268975/>
  22. Zhang Y, Shi W, Zhang W, Mitchison D. Mechanisms of Pyrazinamide Action and Resistance. *Microbiol Spectr* [Internet]. 2013 [citado 6 de febrero de 2021];2(4):1-12. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4268777/>
  23. Juréen P, Werngren J, Toro J-C, Hoffner S. Pyrazinamide Resistance and *pncA* Gene Mutations in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* [Internet]. mayo de 2008 [citado 5 de septiembre de 2021];52(5):1852-4. Disponible en: <https://journals.asm.org/doi/10.1128/AAC.00110-08>
  24. Sheen P, Ferrer P, Gilman RH, López-Llano J, Fuentes P, Valencia E, et al. Effect of pyrazinamidase activity on pyrazinamide resistance in *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)*. marzo de 2009;89(2):109-13.
  25. Jayaraman M, Rajendra SK, Ramadas K. Structural insight into conformational dynamics of non-active site mutations in KasA: A *Mycobacterium tuberculosis* target protein. *Gene*. 15 de diciembre de 2019;720:144082.
  26. Shi W, Zhang X, Jiang X, Ruan H, Barry CE, Wang H, et al. Pyrazinamide inhibits trans-translation in *Mycobacterium tuberculosis*: a potential mechanism for shortening the duration of tuberculosis chemotherapy. *Science* [Internet]. 16 de septiembre de 2011 [citado 6 de febrero de 2021];333(6049):1630-2. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3502614/>
  27. Shi W, Chen J, Feng J, Cui P, Zhang S, Weng X, et al. Aspartate decarboxylase (PanD) as a new target of pyrazinamide in *Mycobacterium tuberculosis*. *Emerging Microbes & Infections* [Internet]. enero de 2014 [citado 3 de septiembre de 2021];3(1):1-8. Disponible en: <https://www.tandfonline.com/doi/full/10.1038/emi.2014.61>
  28. Sousa SF, Fernandes PA, Ramos MJ. Protein-ligand docking: current status and future challenges. *Proteins*. 1 de octubre de 2006;65(1):15-26.
  29. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* [Internet]. noviembre de 2004 [citado 3 de septiembre de 2021];3(11):935-49. Disponible en: <http://www.nature.com/articles/nrd1549>

30. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* [Internet]. diciembre de 2009 [citado 13 de octubre de 2020];30(16):2785-91. Disponible en: <http://doi.wiley.com/10.1002/jcc.21256>
31. Huey R, Morris GM, Olson AJ, Goodsell DS. A semiempirical free energy force field with charge-based desolvation. *J Comput Chem*. 30 de abril de 2007;28(6):1145-52.
32. Hewick RM, Lu Z, Wang JH. Proteomics in Drug Discovery. En: Smith RD, Veenstra TD, editores. *Advances in Protein Chemistry* [Internet]. Academic Press; 2003 [citado 6 de febrero de 2021]. p. 309-42. (Proteome Characterization and Proteomics; vol. 65). Disponible en: <https://www.sciencedirect.com/science/article/pii/S0065323303010246>
33. Karplus M, Kuriyan J. Molecular dynamics and protein function. *Proc Natl Acad Sci U S A*. 10 de mayo de 2005;102(19):6679-85.
34. Gillespie DT. Simulation Methods in Systems Biology. En: Bernardo M, Degano P, Zavattaro G, editores. *Formal Methods for Computational Systems Biology*. Berlin, Heidelberg: Springer; 2008. p. 125-67. (Lecture Notes in Computer Science).
35. Gelpi J, Hospital A, Goñi R, Orozco M. Molecular dynamics simulations: advances and applications. *AABC* [Internet]. noviembre de 2015 [citado 3 de septiembre de 2021];37. Disponible en: <https://www.dovepress.com/molecular-dynamics-simulations-advances-and-applications-peer-reviewed-article-AABC>
36. Bonetta R, Ebejer J-P, Seychell B, Vella M, Hunter T, Hunter GJ. Role of protein structure in drug discovery. diciembre de 2016 [citado 6 de febrero de 2021]; Disponible en: <https://www.um.edu.mt/library/oar/handle/123456789/16001>
37. Liu X, Shi D, Zhou S, Liu H, Liu H, Yao X. Molecular dynamics simulations and novel drug discovery. *Expert Opin Drug Discov*. enero de 2018;13(1):23-37.
38. McCammon JA, Harvey SC. *Dynamics of Proteins and Nucleic Acids* [Internet]. 1.<sup>a</sup> ed. Cambridge University Press; 1987 [citado 3 de septiembre de 2021]. Disponible en: <https://www.cambridge.org/core/product/identifier/9781139167864/type/book>
39. Najmanovich RJ, Torrance JW, Thornton JM. Prediction of Protein Function from Structure: Insights from Methods for the Detection of Local Structural Similarities. *BioTechniques* [Internet]. 1 de junio de 2005 [citado 6 de febrero

de 2021];38(6):847-51. Disponible en: <https://www.future-science.com/doi/full/10.2144/05386TE01>

40. Aono A, Chikamatsu K, Yamada H, Kato T, Mitarai S. Association between *pncA* Gene Mutations, Pyrazinamidase Activity, and Pyrazinamide Susceptibility Testing in Mycobacterium tuberculosis. *Antimicrob Agents Chemother* [Internet]. agosto de 2014 [citado 5 de septiembre de 2021];58(8):4928-30. Disponible en: <https://journals.asm.org/doi/10.1128/AAC.02394-14>
41. Petrella S, Gelus-Ziental N, Maudry A, Laurans C, Boudjelloul R, Sougakoff W. Crystal Structure of the Pyrazinamidase of Mycobacterium tuberculosis: Insights into Natural and Acquired Resistance to Pyrazinamide. *PLOS ONE* [Internet]. 24 de enero de 2011 [citado 5 de enero de 2021];6(1):e15785. Disponible en: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0015785>
42. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinformatics*. 20 de junio de 2016;54:5.6.1-5.6.37.
43. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. *J Cheminform* [Internet]. diciembre de 2011 [citado 3 de septiembre de 2021];3(1):33. Disponible en: <https://jcheminf.biomedcentral.com/articles/10.1186/1758-2946-3-33>
44. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry* [Internet]. 1998 [citado 6 de febrero de 2021];19(14):1639-62. Disponible en: <https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291096-987X%2819981115%2919%3A14%3C1639%3A%3AAID-JCC10%3E3.0.CO%3B2-B>
45. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* [Internet]. septiembre de 2015 [citado 12 de octubre de 2020];1-2:19-25. Disponible en: <https://linkinghub.elsevier.com/retrieve/pii/S2352711015000059>
46. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* [Internet]. 2 de septiembre de 1995 [citado 6 de febrero de 2021];91(1):43-56. Disponible en: <https://www.sciencedirect.com/science/article/pii/001046559500042E>
47. Smith RHB, Dar AC, Schlessinger A. PyVOL: a PyMOL plugin for visualization, comparison, and volume calculation of drug-binding sites

- [Internet]. Pharmacology and Toxicology; 2019 oct [citado 3 de septiembre de 2021]. Disponible en: <http://biorxiv.org/lookup/doi/10.1101/816702>
48. Walker JM, editor. The Proteomics Protocols Handbook [Internet]. Totowa, NJ: Humana Press; 2005 [citado 4 de septiembre de 2021]. Disponible en: <http://link.springer.com/10.1385/1592598900>
  49. RPubS - Matriz de Confusión - Evaluación de modelos de predicción [Internet]. [citado 2 de mayo de 2021]. Disponible en: <https://rpubs.com/chzelada/275494>
  50. Caswell TA, Droettboom M, Lee A, Hunter J, Andrade ESD, Firing E, et al. matplotlib/matplotlib: REL: v3.3.2 [Internet]. Zenodo; 2020 [citado 14 de octubre de 2020]. Disponible en: <https://zenodo.org/record/592536>
  51. Ikai A. Thermostability and aliphatic index of globular proteins. J Biochem. diciembre de 1980;88(6):1895-8.
  52. Weinkam P, Chen YC, Pons J, Sali A. Impact of mutations on the allosteric conformational equilibrium. J Mol Biol [Internet]. 8 de febrero de 2013 [citado 5 de enero de 2021];425(3):647-61. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3557769/>
  53. Lu S, Qiu Y, Ni D, He X, Pu J, Zhang J. Emergence of allosteric drug-resistance mutations: new challenges for allosteric drug discovery. Drug Discovery Today [Internet]. 1 de enero de 2020 [citado 5 de enero de 2021];25(1):177-84. Disponible en: <http://www.sciencedirect.com/science/article/pii/S1359644619303897>
  54. Harmon RC. Pyrazinamide. En: Enna SJ, Bylund DB, editores. xPharm: The Comprehensive Pharmacology Reference [Internet]. New York: Elsevier; 2007 [citado 5 de enero de 2021]. p. 1-5. Disponible en: <http://www.sciencedirect.com/science/article/pii/B9780080552323624924>
  55. Yang L-W, Matysiak S, Hsu S-TD, Mustata Wilson G, Joti Y. Functional dynamics of proteins. Comput Math Methods Med. 2012;2012:242903.
  56. Allana S, Shashkina E, Mathema B, Bablshvili N, Tukvadze N, Shah NS, et al. *pncA* Gene Mutations Associated with Pyrazinamide Resistance in Drug-Resistant Tuberculosis, South Africa and Georgia. Emerg Infect Dis [Internet]. marzo de 2017 [citado 5 de septiembre de 2021];23(3):491-5. Disponible en: [http://wwwnc.cdc.gov/eid/article/23/3/16-1034\\_article.htm](http://wwwnc.cdc.gov/eid/article/23/3/16-1034_article.htm)
  57. Salazar-Salinas K, Baldera-Aguayo PA, Encomendero-Risco JJ, Orihuela M, Sheen P, Seminario JM, et al. Metal-ion effects on the polarization of metal-bound water and infrared vibrational modes of the coordinated metal center of Mycobacterium tuberculosis pyrazinamidase via quantum mechanical calculations. J Phys Chem B. 28 de agosto de 2014;118(34):10065-75.

58. Verlet L. Computer «Experiments» on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys Rev* [Internet]. 5 de julio de 1967 [citado 4 de septiembre de 2021];159(1):98-103. Disponible en: <https://link.aps.org/doi/10.1103/PhysRev.159.98>
59. Bahar I, Lezon TR, Bakan A, Shrivastava IH. Normal Mode Analysis of Biomolecular Structures: Functional Mechanisms of Membrane Proteins. *Chem Rev* [Internet]. 10 de marzo de 2010 [citado 29 de abril de 2021];110(3):1463-97. Disponible en: <https://pubs.acs.org/doi/10.1021/cr900095e>
60. Quan L, Lv Q, Zhang Y. STRUM: structure-based prediction of protein stability changes upon single-point mutation. *Bioinformatics*. 1 de octubre de 2016;32(19):2936-46.
61. Dehury B, Raina V, Misra N, Suar M. Effect of mutation on structure, function and dynamics of receptor binding domain of human SARS-CoV-2 with host cell receptor ACE2: a molecular dynamics simulations study. *Journal of Biomolecular Structure and Dynamics* [Internet]. 7 de agosto de 2020 [citado 29 de abril de 2021];1-15. Disponible en: <https://www.tandfonline.com/doi/full/10.1080/07391102.2020.1802348>
62. El-Bastawissy E, Knaggs MH, Gilbert IH. Molecular dynamics simulations of wild-type and point mutation human prion protein at normal and elevated temperature. *Journal of Molecular Graphics and Modelling* [Internet]. enero de 2001 [citado 29 de abril de 2021];20(2):145-54. Disponible en: <https://linkinghub.elsevier.com/retrieve/pii/S1093326301001139>
63. Supo-Escalante RR, Médico A, Gushiken E, Olivos-Ramírez GE, Quispe Y, Torres F, et al. Prediction of Mycobacterium tuberculosis pyrazinamidase function based on structural stability, physicochemical and geometrical descriptors. Silman I, editor. *PLoS ONE* [Internet]. 31 de julio de 2020 [citado 29 de abril de 2021];15(7):e0235643. Disponible en: <https://dx.plos.org/10.1371/journal.pone.0235643>
64. Wang Y, Harrison CB, Schulten K, McCammon JA. Implementation of accelerated molecular dynamics in NAMD. *Comput Sci Disc* [Internet]. 1 de marzo de 2011 [citado 20 de septiembre de 2021];4(1):015002. Disponible en: <https://iopscience.iop.org/article/10.1088/1749-4699/4/1/015002>
65. Bowerman S, Wereszczynski J. Detecting Allosteric Networks Using Molecular Dynamics Simulation. En: *Methods in Enzymology* [Internet]. Elsevier; 2016 [citado 29 de abril de 2021]. p. 429-47. Disponible en: <https://linkinghub.elsevier.com/retrieve/pii/S0076687916300623>
66. Egorov AM, Ulyashova MM, Rubtsova MY. Bacterial Enzymes and Antibiotic Resistance. *Acta Naturae*. diciembre de 2018;10(4):33-48.

67. Yang J, Nune M, Zong Y, Zhou L, Liu Q. Close and Allosteric Opening of the Polypeptide-Binding Site in a Human Hsp70 Chaperone BiP. Structure [Internet]. diciembre de 2015 [citado 4 de septiembre de 2021];23(12):2191-203. Disponible en: <https://linkinghub.elsevier.com/retrieve/pii/S0969212615004517>
68. Ranaghan KE, Lonsdale R. QM/MM Methods. En: Roberts GCK, editor. Encyclopedia of Biophysics [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013 [citado 29 de abril de 2021]. p. 2154-6. Disponible en: [http://link.springer.com/10.1007/978-3-642-16712-6\\_275](http://link.springer.com/10.1007/978-3-642-16712-6_275)
69. Feng C, Wang H, Lu N, Chen T, He H, Lu Y, et al. Log-transformation and its implications for data analysis. Shanghai Arch Psychiatry. abril de 2014;26(2):105-9.
70. Ramirez-Busby SM, Valafar F. Systematic Review of Mutations in Pyrazinamidase Associated with Pyrazinamide Resistance in Mycobacterium tuberculosis Clinical Isolates. Antimicrob Agents Chemother [Internet]. septiembre de 2015 [citado 20 de septiembre de 2021];59(9):5267-77. Disponible en: <https://journals.asm.org/doi/10.1128/AAC.00204-15>

## IX. ANEXOS

**Anexo 1.** *Bash script* que realiza mutaciones puntuales usando Modeller.

```
#####
# By: Gustavo E. Olivos-Ramirez #
# gustavo.olivos@upch.pe #
# Lima-Peru #
#####

#AA code
A=ALA;C=CYS;D=ASP;E=GLU;F=PHE;G=GLY;H=HIS;I=ILE;K=LYS;L=LEU;M=MET;
N=ASN;P=PRO;Q=GLN;R=ARG;S=SER;T=THR;V=VAL;W=TRP;Y=TYR

#list of mutations
mutant=("344 $S" "367 $F" "382 $L" "439 $K" "477 $N" "478 $I" "479 $S" "520 $S" "522 $V")

# mutate with modeller
for m in "${mutant[@]"; do
python mutate.py RBD ${m} E > mutate.log
i=${m/ /}
mkdir -pv ${i}
mv RBD*.pdb "${i}"/
mv *.log "${i}"/
mv "${i}"/RBD.pdb .
echo "${m} mutation done";
done
```

**Anexo 2.** *Bash script* que convierte estructuras \*.SDF a \*.PDB

```
#####
# ----- This is a script for convert library ligands of sdf to pdb ----- #
# AUTOR: ROPÓN-PALACIOS G., BSc, MSc(c), #
# Bioinformatics & Theoretical Biophysics - Molecular Genetics Lab, #
# Departament of Biology & Physics, Faculty of Sciences, #
# Universidad Nacional de San Antonio Abad del Cusco #
# Av. La Cultura 733, Wanchaq, Cusco, Peru. #
# E-mail: biodano.geo@gmail.com #
#####

# Run script usign babel module of Openbabel:
for i in example_*.sdf; do
b=`basename $i .sdf`
echo Processing ligando $b
mkdir -p $b
babel -isdf $i -opdb ${b}/out.pdb --gen3D -p 7.4;
done

# -isdf is a option for input in sdf format of ligand
# -opdb is a option for output in pdb format
# --gen3D is a option for generate 3D coordinates
# -p 7.4 is a option for add Hydrogens to pH 7.4

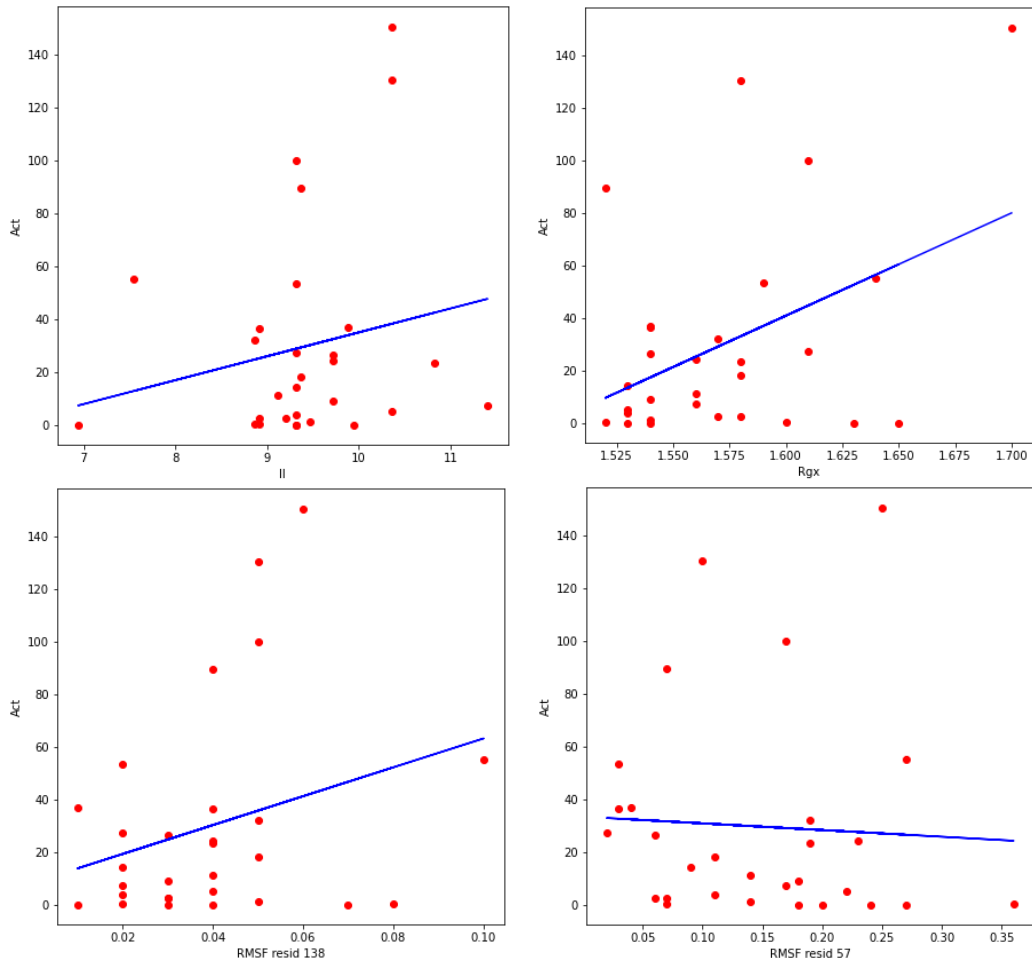
# REMEMBER:
# If you are usign this script please, cited link of github
# For more information o problems please send me a e-mail.
```

### Anexo 3. TCL script que determina los SASA.

```
#####
# sasa.tcl #
# DESCRIPTION: #
# This script is quick and easy to provide procedure #
# for computing the Solvent Accessible Surface Area (SASA) #
# of Protein and allows Users to select regions of protein. #
# #
# EXAMPLE USAGE: #
# source sasa.tcl #
# Selection: chain A and resid 1 #
# #
# AUTHORS: #
# Sajad Falsafi (sajad.falsafi@yahoo.com) #
# Zahra Karimi #
# 3 Sep 2011 #
#####

puts -nonewline "\n \t \t Selection: "
gets stdin selmode
# selection
set sel [atomselect top "$selmode"]
set protein [atomselect top "protein"]
set n [molinfo top get numframes]
set output [open "SASA_$selmode.dat" w]
# sasa calculation loop
for {set i 0} {$i < $n} {incr i} {
    molinfo top set frame $i
    set sasa [measure sasa 1.4 $protein -restrict $sel]
    puts "\t \t progress: $i/$n"
    puts $output "$sasa"
}
puts "\t \t progress: $n/$n"
puts "Done."
puts "output file: SASA_$selmode.dat"
close $output
```

**Anexo 4.** Líneas de tendencias de las principales variables consideradas en el modelo de regresión lineal múltiple.



**Anexo 5.** Análisis de regresión lineal múltiple en Epi Info.

**Linear Regression**

Variable	Coefficient	Std Error	F-test	P-Value
II	23.388	7.953	8.6482	0.007140
Rgx	400.849	169.875	5.5680	0.026764
RMSF138	1156.166	471.770	6.0059	0.021924
RMSF57	-252.103	90.251	7.8027	0.010084
CONSTANT	-828.391	270.577	9.3732	0.005359

Correlation Coefficient:  $r^2 = 0.48$

Source	df	Sum of Squares	Mean Square	F-statistic
Regression	4	21558.398	5389.599	5.469
Residuals	24	23653.250	985.552	
Total	28	45211.648		

**REGRESS Act = Rgx RMSF138 RMSF57 Rgx \* RMSF138 RMSF138 \* RMSF57 Rgx \* RMSF57**

**Anexo 6.** Varianza de los parámetros cinéticos.

Mutación	<i>Kcat</i>	<i>Km</i>	<i>Efficiency</i>	<i>Activity</i>
WT	395.350814	1259.105	321.380965	2286.03516
A102V	348.494451	854.431	41.0196895	134.004741
A171T	126.00995	1754.019	104.141885	148.076198
A46V	58.5447538	1654.482	377.258218	144.280079
C14G	0.12678844	428.557	0.02518277	1.46394056
D12A	3.64683452	166.493	6.17275442	69.0110948
D12G	2.21399577	23.413	64.5008521	43.0664063
D136G	74.2249575	4267.04	81.6843056	317.285156
D49N	0.00200327	9209.807	3.84E-05	0.00132921
F58L	58.7146055	3544.762	157.196756	791.856669
F94L	122.899942	1436.655	14.2795161	535.97073
G24D	2.62838396	406.478	177.29322	71.6315375
G78C	0.48751482	127.471	0.57193047	4.67190213
H51R	9.89E-07	210.718	0.00753147	6.78E-06
H57R	0.02376714	318.133	0.01567929	0.09285495
H71Y	76.4000593	1447393.34	0.09760783	5.33821771
K48T	0.68809072	16.259	24.6294518	7.61990017
L116P	71.7944438	146.332	21.1673004	218.793403
L172P	0.11487285	8860.208	1.12620824	0.04427247
L4S	6.33E-11	124.107	4.03E-08	0.00207689
M175V	124.954505	14251.06	11.4268132	3.3068934
P54L	1.78922853	144.931	1.49370156	83.7415636
P62L	4.64500437	154.661	13.4533345	61.4019165
P62R	6.95376407	291.026	4.47056615	10.5195433
Q10P	0.20657491	164.522	1.16659282	16.6736111
R29P	0.05547569	37.545	100.079981	0.73405626
T135P	0.00061829	1316.988	---	0.00271267
T142A	9.89E-07	1533.417	6.21E-07	4.12E-06
T160K	0.12537577	1333466.19	1.54E-06	0.00035156
T76P	0.60187015	16.259	123.698741	6.12046984
V125F	279.619661	5962.07	305.820044	68.8347376
V139A	20651.5718	22896742.1	0.05317524	25.9736487
V180F	2565.81996	349428.621	407.888032	50.0857097
W119L	0.09753779	679.317	0.03187194	1.08669766
Y34D	436.516155	7170.265	2014.30254	472.835965
Y64D	898.192691	1327.251	112.518806	4423.63553

**Anexo 7.** Parámetros cinéticos transformados.

Mutación	<i>K<sub>cat</sub></i>	<i>K<sub>m</sub></i>	<i>Efficiency</i>	<i>Activity</i>
WT	1.992	1.974	1.993	1.955
A102V	0.703	2.094	1.928	1.948
A171T	0.613	2.089	1.63	1.766
A46V	0.539	1.949	1.559	1.6
C14G	0.188	2.021	0.369	0.32
D12A	0.663	1.897	1.488	1.357
D12G	0.769	1.644	1.947	1.555
D136G	0.643	2.462	1.16	1.449
D49N	-0.412	2.367	-1.168	-0.951
F58L	0.562	1.959	1.504	1.466
F94L	0.707	2.196	1.634	1.707
G24D	0.448	1.464	1.426	0.948
G78C	0.5	1.932	1.082	1.255
H51R	-0.886	2.052	-2.6	-1.812
H57R	0.035	1.939	0.131	0.108
H71Y	-0.119	2.699	0.047	0.498
K48T	0.684	1.547	1.834	1.433
L116P	0.965	2.098	2.021	2.113
L172P	0.085	2.006	0.151	0.255
L4S	-2.127	1.124	-3.524	-1.595
M175V	0.341	2.22	1.163	1.319
P54L	0.548	1.902	1.241	1.085
P62L	0.609	1.79	1.509	1.403
P62R	0.674	2.098	1.337	1.453
Q10P	0.275	1.707	0.851	0.606
R29P	0.382	1.235	1.494	0.854
T135P	-0.678	1.829	---	-1.434
T142A	-1.2	1.923	-2.31	-2.295
T160K	-0.296	2.88	-1.492	-1.009
T76P	0.646	1.392	1.902	1.367
V125F	0.637	2.105	1.782	1.857
V139A	-0.362	2.907	0.485	0.894
V180F	0.097	2.565	1.239	1.508
W119L	0.103	1.983	0.232	0.321
Y34D	0.246	1.618	1.652	1.696
Y64D	0.623	2.079	2.034	2.139

## X. APÉNDICE

### **Apéndice A.** Repositorio de scripts.

Los principales algoritmos desarrollados en esta tesis pueden ser obtenidos en los repositorios de Git-Hub, mediante los siguientes enlaces:

<https://github.com/bioinfo-upch>

<https://github.com/tavolivos>

### **Apéndice B.** Resultados publicados.

Los resultados de los modelos obtenidos mediante las transformaciones logarítmicas y las dinámicas moleculares realizadas en esta tesis fueron publicados en el siguiente artículo:

Supo-Escalante RR, Médico A, Gushiken E, Olivos-Ramírez GE, Quispe Y, ...  
Mirko Zimic (2020) Prediction of Mycobacterium tuberculosis pyrazinamidase function based on structural stability, physicochemical and geometrical descriptors. PLOS ONE 15(7): e0235643. <https://doi.org/10.1371/journal.pone.0235643>

## RESEARCH ARTICLE

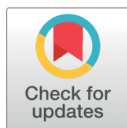
# Prediction of *Mycobacterium tuberculosis* pyrazinamidase function based on structural stability, physicochemical and geometrical descriptors

Rydberg Roman Supo-Escalante<sup>1</sup>, Aldhair Médico<sup>1</sup>, Eduardo Gushiken<sup>1</sup>, Gustavo E. Olivos-Ramírez<sup>1</sup>, Yaneth Quispe<sup>1</sup>, Fiorella Torres<sup>1</sup>, Melissa Zamudio<sup>1</sup>, Ricardo Antiparra<sup>1</sup>, L. Mario Amzel<sup>2</sup>, Robert H. Gilman<sup>3</sup>, Patricia Sheen<sup>1</sup>, Mirko Zimic<sup>1\*</sup>

**1** Laboratorio de Bioinformática, Biología Molecular y Desarrollos Tecnológicos, Facultad de Ciencias y Filosofía, Universidad Peruana Cayetano Heredia, Lima, Peru, **2** Department of Biophysics and Biophysical Chemistry, Johns Hopkins University, Baltimore, MD, United States of America, **3** International Health Department, Johns Hopkins School of Public Health, Baltimore, MD, United States of America

 These authors contributed equally to this work.

\* [mirko.zimic@upch.pe](mailto:mirko.zimic@upch.pe)



## OPEN ACCESS

**Citation:** Supo-Escalante RR, Médico A, Gushiken E, Olivos-Ramírez GE, Quispe Y, Torres F, et al. (2020) Prediction of *Mycobacterium tuberculosis* pyrazinamidase function based on structural stability, physicochemical and geometrical descriptors. PLoS ONE 15(7): e0235643. <https://doi.org/10.1371/journal.pone.0235643>

**Editor:** Israel Silman, Weizmann Institute of Science, ISRAEL

**Received:** March 30, 2020

**Accepted:** June 19, 2020

**Published:** July 31, 2020

**Copyright:** © 2020 Supo-Escalante et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The data underlying the results presented in the study are available in the manuscript and in the supplementary material.

**Funding:** This research was funded by the Wellcome Trust Intermediate Fellowship (grant 099805/Z/12/Z) (PS). This study was also partially funded Grand Challenge Canada (grant 0687-01-10) (PS), by FONDECYT-Peru (grant 037-2014) (MZ). The funders had no role in study design, data

## Abstract

### Background

Pyrazinamide is an important drug against the latent stage of tuberculosis and is used in both first- and second-line treatment regimens. Pyrazinamide-susceptibility test usually takes a week to have a diagnosis to guide initial therapy, implying a delay in receiving appropriate therapy. The continued increase in multi-drug resistant tuberculosis and the prevalence of pyrazinamide resistance in several countries makes the development of assays for prompt identification of resistance necessary. The main cause of pyrazinamide resistance is the impairment of pyrazinamidase function attributed to mutations in the promoter and/or *pncA* coding gene. However, not all *pncA* mutations necessarily affect the pyrazinamidase function.

### Objective

To develop a methodology to predict pyrazinamidase function from detected mutations in the *pncA* gene.

### Methods

We measured the catalytic constant ( $k_{cat}$ ),  $K_M$ , enzymatic efficiency, and enzymatic activity of 35 recombinant mutated pyrazinamidase and the wild type (Protein Data Bank ID = 3pl1). From all the 3D modeled structures, we extracted several predictors based on three categories: structural stability (estimated by normal mode analysis and molecular dynamics), physicochemical, and geometrical characteristics. We used a stepwise Akaike's information criterion forward multiple log-linear regression to model each kinetic parameter with each

collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

category of predictors. We also developed weighted models combining the three categories of predictive models for each kinetic parameter. We tested the robustness of the predictive ability of each model by 6-fold cross-validation against random models.

## Results

The stability, physicochemical, and geometrical descriptors explained most of the variability ( $R^2$ ) of the kinetic parameters. Our models are best suited to predict  $k_{cat}$ , efficiency, and activity based on the root-mean-square error of prediction of the 6-fold cross-validation.

## Conclusions

This study shows a quick approach to predict the pyrazinamidase function only from the *pncA* sequence when point mutations are present. This can be an important tool to detect pyrazinamide resistance.

## Introduction

Tuberculosis (TB) is a major infectious disease caused by *Mycobacterium tuberculosis* (MTB) which mostly affects people in developing countries. According to the WHO's global TB report of 2018, TB is one of the top 10 causes of death and in 2017 there were about 1.6 million TB deaths and 10 million infections [1]. Emerging drug resistance hinders the progress and efforts to control this disease [2]. Among all drugs available, pyrazinamide (PZA) is an important anti-tuberculosis drug against the dormant or semi-dormant latent stage of MTB [3]. Despite its importance, PZA is also responsible for a relevant proportion of treatment abandons because of side effects [4]. In Peru, over 30% of multidrug-resistant (MDR) TB strains are also resistant to PZA [5]. Moreover, in other countries exists a large prevalence of PZA resistance [6].

The mechanism of action and resistance to PZA in MTB is not entirely understood [3]. PZA, the pro-drug, enters the cytoplasm of MTB by passive diffusion and is hydrolyzed into pyrazinoic acid (POA) by the enzyme pyrazinamidase (PZAse), encoded by the *pncA* gene [7]. POA, the active drug, accumulates in the cytoplasm and enters a cycle of entry and exit from MTB aided by an efflux pump not yet identified. Outside, in an acidic environment, POA is protonated and when re-enters the cytoplasm, releases the protons and causes membrane disruption and cellular damage [8].

The major mechanism of PZA resistance, according to several studies, is the loss of PZAse activity due to mutations in the *pncA* gene, suppressing its ability to hydrolyze PZA [9–12]. Mutations directly affecting the active site (AS) (Cys138, Asp8, and Lys96), or the metal coordination site (MCS) (Asp49, His51, His57, and His71) are more likely to impair PZAse function [13]. Current knowledge of *pncA* gene sequences has shown that PZA resistant strains are associated with *pncA* mutations scattered throughout the entire gene which deplete the PZAse function [10,14–16]. It is important to highlight that a failure in the PZAse function strictly causes resistance to PZA.

The current gold standard test to detect PZA resistance is the colorimetric MGIT 960-TB liquid culture. Alternate assays comprise the MODS [17] and MODS-PZA [18] methods, based on microscopic observation and the colorimetric Wayne test and variants like the reported by Aono *et al.*, 2018 [19] or Alcántara *et al.*, 2019 [20]. The latter detect expelled POA

in the extracellular environment, a biomarker of PZA resistance [14,21], and indirectly measures the PZAse activity [20]. Nevertheless, the high cost, duration, or lack of reproducibility of these tests limit their use. Therefore, an accurate, simple, and fast test to determine PZA resistance is required to prevent the unnecessary use of PZA during anti-tuberculous treatment.

Given the current extensive use of next-generation sequencing (NGS), it has been proposed as an alternative to detect PZA-resistance by the detection of mutations in the *pncA* gene [22–24]. Nowadays, the sequencing of this gene is more affordable, numerous studies have reported *pncA* polymorphisms associated with PZA resistance with high accuracy [10,22,25–27]. Nevertheless, most of them are focused on susceptible/resistant binary prediction instead of predicting PZAse function and PZA resistance level. The main limitation of this approach is the fact that not every *pncA* mutation impairs the PZAse enzymatic function. To overcome this limitation, a method to predict the enzymatic function based only on the *pncA* sequence is needed.

Several studies have addressed the prediction of the enzymatic function of enzymes based on the predicted structure of mutated enzymes [28–34]. The approach of these studies used several features as predictors, including electrostatic potentials used in quantitative structure-activity relationships (QSAR) [35–37] and geometrical descriptors that are measurements of distances between specific points of interest and indirectly involve short and long-range interactions between residues [38–40].

Amino acidic mutations affecting protein function may alter the protein structural stability [31,41,42]. In particular, structural stability has been correlated with the enzymatic function in PZAse [43]. This property could be studied by several techniques like molecular dynamics (MD) simulations, using the root-mean-square fluctuation (RMSF) of residues, to find unstable regions with a high degree of movement and flexibility [44,45]. In PZAse the most unstable region is the flap region, a loop from His51 to His71 [45–49], which is the lid of the PZA/ Fe<sup>2+</sup> binding cavity [13], and alterations of its stability have been reported in mutated PZAases associated with PZA resistance [46–49].

An alternative, faster and reliable method to analyze protein stability, is the Normal Mode Analysis (NMA), which is based on the protein structure [50,51]. NMA considers oscillating movements that describe relevant motions of small amplitude at the atomic or aminoacidic level [52,53]. It assumes that proteins oscillate harmonically around a given conformation [54] and calculates the normal modes of vibrations to describe the overall motion. From this, fluctuation scores for each position in the protein are calculated [55].

In this study, we developed statistical models to predict PZAse function, based on structural stability, physicochemical, and geometrical descriptors inferred from 3D modeled structures of PZAse based only on the mutated sequence. Multiple log-linear regressions and 6-fold cross-validation were used to get the best linear predictors and to verify robustness.

## Methods

### Cloning and expression of PZAse

The sequences of 35 *pncA* mutant genes from clinical strains, among the H37Rv wild-type were cloned and expressed in *Escherichia coli* LEMO21 as reported before [7]. Briefly, each *pncA* gene sequence was amplified by PCR from genomic DNA using primers containing restriction sites for *NcoI* and *XhoI* and inserted into pET28a plasmid containing a COOH-terminal 6-His tag. Both PCR product and pET28a plasmid were digested with *NcoI* and *XhoI*, ligated and transformed into *E. coli* LEMO21. Proteins were purified by affinity chromatography using a HisTrap HP column (Novagen).

### Kinetic parameters of recombinant PZAses

PZase kinetic parameters were calculated using the hydrolysis reaction of PZA. Briefly, PZA was used from 0 to 5 mM and incubated with 1  $\mu$ M PZase in 50 mM sodium phosphate pH 6.5. To prevent the hydrolysis of more than 10% of the initial PZA, an incubation period of 1 min was used. It was increased to 2 h for mutants with very low activity. 10  $\mu$ L of 20% Ferrous ammonium sulfate was added, followed immediately by adding 445  $\mu$ L of 0.2 mM glycine-HCl (pH 3.4) to stop the reaction. Precipitates were removed by centrifugation (11,000 rpm for 5 min). Absorbance (OD) was measured at 450 nm in a 96-well plate using 200  $\mu$ L of reaction. The amount of POA produced was estimated by interpolation in a standard curve of known concentrations. Each recombinant PZase was tested at least 5 times, and at least 2 different groups of recombinant proteins were analyzed for each strain. Finally, we estimated  $k_{cat}$ ,  $K_M$ , efficiency, and activity as before [7].

To make the measurements of the kinetic parameters comparable between the two batches of recombinant proteins, we normalized each kinetic parameter by dividing its value by the one of the wild type (WT) H37RV PZase of the corresponding batch.

### Theoretical structural modeling of mutated PZAses

All mutants were modeled using the protein modeling server SWISS-MODEL [56] with the crystal structure of the WT PZase available in Protein Data Bank (PDB ID: 3PL1, 2.2  $\text{\AA}$  resolution) as a template.

### Stability analysis of amino-acidic fluctuations by NMA

We evaluated structural stability using the fluctuation scores for each amino acid position calculated by NMA [57]. Briefly, a potential energy function at residue level is constructed for each structure using an elastic network model that connects all the  $C\alpha$ 's of residues that interact with each other with a spring. The original code was modified to consider connected the  $C\alpha$ 's of pairs of residues for which the distance between any main chain or side chain atoms was less than 5  $\text{\AA}$ . All  $C\alpha$ - $C\alpha$  pairwise forces were modeled as a harmonic oscillator with the same spring constant. Then, the modes are calculated by solving the following equation:

$$HX = DTX \quad (1)$$

Where H is the Hessian matrix of the potential energy function, T the kinetic energy, X the eigenvectors, and D a diagonal matrix containing the eigenvalues [57]. The frequency of oscillation of each normal mode is related to the eigenvalues and is the same for all the residues in the respective normal mode. Each normal mode has an associated eigenvector of length equal to the number of residues (or 3N). The components of the eigenvector corresponding to the amplitude and the direction of movement of each  $C\alpha$ .

We considered 40 modes, from the 7<sup>th</sup> up to the 46<sup>th</sup> eigenvalues. The first six were left out because they correspond to the 3 translations and 3 rotations and were almost zero. After that, the fluctuations for each amino acid in each mutated structure were calculated by the equation [55]:

$$\langle x_i^2 \rangle = \frac{k_B T}{m_i} \sum_{j=i}^n \frac{a_{ij}^2}{w_j^2} \quad (2)$$

Where  $k_B$  is the Boltzmann constant, T the absolute temperature,  $m_i$  the mass of residue i, n the number of modes,  $a_{ij}$  the eigenvector for residue i in mode j, and  $w_j$  the square root of the eigenvalue for mode j. A final dataset of 185 variables for each structure was produced.

### Stability analysis of amino-acidic RMSF by MD

As an alternative approach to stability calculated by NMA, molecular dynamics (MD) simulations were performed on each PZAse structure. For this purpose, the software GROMACS version 2018.3 [58] in GPU configuration was used to simulate the behavior in a solvent of the mutated structures modeled with SWISS-MODEL [56] and the WT-PZAse (PDB ID: 3PL1) [13].

During the simulation, topologies were generated with the PDB2GMX module, using the OPLS-AA/L all-atom force field [59]. Additionally, each structure was centered in a 10 Å cubic box, with periodic boundary conditions. Water molecules from the SPC/E model [60] were used for solvation, and charges were neutralized by adding Na<sup>+</sup> and Cl<sup>-</sup> ions.

Afterward, energy minimization of the system was performed using the steepest descent algorithm with 50 000 steps until obtaining energy lower than 1000 Kcal mol<sup>-1</sup>. The temperature and the pressure of the system were equilibrated by using NVT and NPT assembling, respectively. In the first stage, the temperature was equilibrated to 310.15 K and constantly maintained with the Berendsen thermostat [61] for 2 ns. In the second stage, the system's pressure was equilibrated to 1 bar and constantly maintained with the Parrinello-Rahman barostat [62,63] for 4 ns. Finally, the simulation time was 500 ns, with 2 fs integration step and constant pressure and temperature conditions, using the integration algorithm leap-frog [64]. For generating the trajectories, the LINCS algorithm [65] was used to restrict interactions during equilibrium, while the Particle-Mesh Ewald algorithm [66] was used to restrict the long-range ionic interactions.

The trajectories obtained from each simulation were centered and used to calculate the root-mean-square deviation (RMSD) of the protein of each structure. Likewise, the trajectories corresponding to the last 100 ns of each molecular dynamics (400 to 500 ns) were extracted to calculate the root-mean-square-fluctuations (RMSF) of the backbone. The data was extracted in \*.XGV format and RMSD and RMSF plots were generated using an *in-house* python script using the Matplotlib library [67]. A final dataset of 185 RMSF variables for each structure was generated.

### Calculation of physicochemical descriptors

The physicochemical descriptors evaluated were the differences in electrostatic potential (EP) between residues from each mutated protein and residues from the WT-PZAse and the sum of all differences in electrostatic potential in the entire protein. The MutantElec server [68] was used to calculate EPs, this server uses an Adaptive Poisson-Boltzmann Solver (APBS) to analyze the effect of solvents in proteins and as input requires the dielectric constant of protein ( $\epsilon_{\text{protein}}$ ) and water ( $\epsilon_{\text{water}}$ ) at a fixed temperature. We assumed a value of  $\epsilon_{\text{protein}}$  of 4 as a mean for proteins and calculated  $\epsilon_{\text{water}} = 74.1522$  at  $T = 37^\circ\text{C}$  (which represents lung temperature) as previously reported [69]. A final dataset of 185 physicochemical variables for each structure was generated.

### Calculation of geometrical descriptors for WT and mutated PZAse

For every mutated PZAse structure, we calculated 4 references as described previously [38] (S2 Fig): (i) Point B: the barycenter of the AS (Asp8, Lys96, and Cys138) and the MCS (Asp49, His51, and His71 excluding His57). (ii) Point P: the point of projection of the C $\alpha$  of the residues on the plane formed by the trio of amino acids (AS/MCS). (iii) Point I: the point of intersection between the resulting vector (V<sub>r</sub>) of each amino acid regarding the C $\alpha$  and the plane formed by the trio of amino acids (AS/MCS). (iv) Point T: the barycenter of the MCS including His57.

Using the points of interest, 24 different descriptors were calculated on each of the 185 aminoacidic positions of PZAse resulting in 4440 variables. The exact meaning of each geometrical descriptor is available in the [S1 Appendix](#).

### Construction and validation of predictive linear models

Since kinetic parameters of enzymes are strictly positive variables and tend to follow a log-normal distribution in nature [70], we assumed them as log-normally distributed. The following transformation was used to go from the mean of the kinetic parameters to the mean of the logarithmic kinetic parameters:

$$\overline{\log_{10}(x)} = \log_{10}\left(\frac{\bar{x}^2}{\sqrt{\bar{x}^2 + \sigma_x^2}}\right) \quad (3)$$

Where  $\bar{x}$  refers to the mean value of the respective kinetic parameter, and  $\sigma_x^2$  to its variance. We worked with these estimates to construct the log-linear models. Each kinetic parameter was combined with the stability, physicochemical and geometrical datasets separately. Observations with a missing value in the respective kinetic parameter were removed from each dataset. Then, covariates with missing values or variance equal to zero were dropped.

To avoid correlation between covariates, simple log-linear regressions between the kinetic parameter and the rest of covariates were performed for each dataset. The p-value for each regression was used to sort covariates in ascending order. Then, we went through the sorted list and calculated Pearson's correlation coefficient between the top covariate and the rest of the covariates under it. If the correlation coefficient was greater than 0.8, the covariates down in the list were removed from the dataset and the procedure continue with the next covariate in the reduced list until the last covariate is reached.

Once the datasets were reduced, covariates were selected using a stepwise Akaike's information criterion (AIC) forward regression until having 10 covariates. The final individual models include the best combination of 6 out of the 10 selected covariates that display the highest adjusted  $R^2$ . The weighted models were constructed for each kinetic parameter using the fitted values of their respective individual models as covariates.

To prevent overfitting, we performed a repeated 6-fold cross-validation for the weighted models and the individual models of stability, physicochemical, and geometrical descriptors using the R's package Caret [71]. The root-mean-square error (RMSE) calculation was used to evaluate the predictive ability of each model. To have a clear idea of the distribution of RMSE values of our models, we performed a repeated 6-fold cross-validation 1000 times and compare the distribution of the RMSE for individual models with the one produced by models with random descriptors (random models). For the weighted model, the RMSE distribution was compared against the one produced by the individual models.

## Results

### Kinetic parameters of recombinant PZAses

The kinetic parameters measured for the 35 mutated and WT PZAses revealed a wide range of variation ([Table 1](#) and [S1 Table](#)). Due to experimental error, the WT H37Rv PZAse showed slightly different estimates of the kinetic parameters between the two batches of proteins produced. To prevent a batch-bias, we expressed all the kinetic parameters as the corresponding percentage of the WT-PZAse within each batch and worked with these new variables defined as Relative- $k_{cat}$ , Relative- $K_M$ , Relative-efficiency, and

Table 1. Kinetic parameters of two batches of mutated PZAses from *M. tuberculosis*.

	Mutant	$k_{cat}$ ( $\text{min}^{-1}$ )	Relative $k_{cat}$ (%)	$K_M$ (mM)	Relative $K_M$ (%)	Efficiency ( $\text{mM}^{-1}\text{min}^{-1}$ )	Relative Efficiency (%)	Activity ( $\mu\text{mol POA min}^{-1}\text{mg}^{-1}$ PZase)	Relative Activity (%)
Batch 1	WT	1005.410	100.000	1.240	100.000	806.600	100.000	38.400	100.000
	A102V	1084.220	107.839	1.581	127.479	685.890	85.035	34.316	89.364
	C14G	25.110	2.497	1.327	107.014	18.920	2.346	0.902	2.350
	<b>D12A</b>	245.170	24.385	0.990	79.839	248.790	30.844	9.240	24.063
	<b>D12G</b>	368.460	36.648	0.550	44.355	716.830	88.871	14.000	36.458
	D136G	490.270	48.763	3.680	296.774	133.030	16.493	12.340	32.135
	<b>D49N</b>	1.530	0.152	3.090	249.194	0.550	0.068	0.045	0.117
	F58L	355.880	35.397	1.297	104.627	274.300	34.007	14.132	36.801
	<b>F94L</b>	712.920	70.908	2.000	161.290	348.600	43.218	21.190	55.182
	<b>G24D</b>	100.130	9.959	0.420	33.871	236.200	29.283	4.280	11.146
	<b>G78C</b>	105.160	10.459	1.070	86.290	97.650	12.106	6.960	18.125
	<b>H51R</b>	0.170	0.017	1.410	113.710	0.120	0.015	0.006	0.016
	H57R	12.020	1.196	1.099	88.609	10.940	1.356	0.505	1.316
	H71Y	69.840	6.946	10.657	859.464	9.320	1.155	1.425	3.710
	<b>K48T</b>	241.820	24.052	0.440	35.484	551.200	68.336	10.450	27.214
	<b>L116P</b>	1324.500	131.737	1.560	125.806	847.660	105.091	50.150	130.599
	L4S	0.001	0.000	0.201	16.173	0.003	0.000	0.015	0.039
	P54L	141.195	14.044	1.001	80.763	140.989	17.479	5.529	14.398
	P62L	203.859	20.276	0.779	62.802	261.782	32.455	10.129	26.378
	Q10P	37.665	3.746	0.650	52.431	57.932	7.182	1.978	5.151
	R29P	59.000	5.868	0.225	18.117	262.881	32.591	2.761	7.189
	<b>T135P</b>	0.450	0.045	0.930	75.000	-	-	0.020	0.052
	T142A	0.040	0.004	1.130	91.090	0.040	0.005	0.002	0.005
	T160K	3.500	0.348	13.630	1099.194	0.260	0.032	0.038	0.100
	<b>T76P</b>	202.980	20.189	0.310	25.000	650.330	80.626	8.990	23.411
	V139A	628.710	62.533	25.422	2050.140	24.730	3.066	3.455	8.997
W119L	17.090	1.700	1.234	99.508	13.850	1.717	0.883	2.299	
<b>Y34D</b>	386.400	38.432	0.830	66.935	460.190	57.053	20.580	53.594	
Y64D	1357.250	134.995	1.548	124.870	876.550	108.672	57.830	150.598	
Batch 2	WT	739.480	100.000	1.162	100.000	648.776	100.000	27.607	100.000
	A171T	408.035	55.179	1.501	129.177	283.986	43.773	16.422	59.484
	A46V	244.662	33.086	1.121	96.436	260.875	40.210	11.433	41.411
	L172P	11.730	1.586	1.469	126.390	10.877	1.677	0.500	1.811
	M175V	199.743	27.011	2.260	194.516	96.922	14.939	5.776	20.921
	P62R	206.273	27.894	1.469	126.430	141.472	21.806	7.889	28.574
	V125F	609.467	82.418	1.678	144.418	407.886	62.870	19.996	72.428
	V180F	474.632	64.185	6.309	542.873	149.531	23.048	9.104	32.975

The relative kinetic parameters are expressed as percentages of the corresponding WT-PZase kinetic parameters. The mutants previously reported by our group (Sheen et al., 2009) are in bold. For the mutant T135P, the experimental measurements of efficiency are not available.

<https://doi.org/10.1371/journal.pone.0235643.t001>

Relative-activity. The relative kinetic parameters revealed a high Pearson correlation coefficient (R) between  $k_{cat}$  and efficiency ( $R = 0.787$ ),  $k_{cat}$  and activity ( $R = 0.947$ ) and efficiency and activity ( $R = 0.865$ ), while  $K_M$ , was neither well correlated with  $k_{cat}$  ( $R = 0.072$ ), efficiency ( $R = -0.286$ ), nor activity ( $R = -0.160$ ).

### Structural modeling of mutated PZAses and stability, physicochemical, and geometrical structural analysis

The modeled structures of the mutated PZAses, showed an average root-mean-square deviation (RMSD) of 0.094 Å, at all-atoms level, compared to the WT-PZase crystal structure (PDB: 3PL1) (S1 Fig).

We estimated, first, 4440 geometrical descriptors of distances between specific points of interest related to the AS/MCS and all the residue positions of the protein (S1 Dataset). A more detailed explanation of the geometrical meaning of these descriptors is available as supplementary material (S2 Fig and S1 Appendix). Second, 185 physicochemical descriptors that include the difference in electrostatic potential per residue (DEPR) against the WT-PZase and the global difference in electrostatic potential (GDEP) (S2 Dataset and S3 Fig). Third, 185 stability descriptors of amino-acidic fluctuations calculated by NMA that describe the degree of fluctuation of each residue averaged over 40 normal modes of vibration (S3 Dataset and S4 Fig). Alternatively, we performed MD simulations of 500 ns (S5 Fig) on each structure to calculate the RMSF of the protein backbone and to use it as an alternative descriptor of stability for each position; other 185 descriptors were generated (S4 Dataset and S6 Fig).

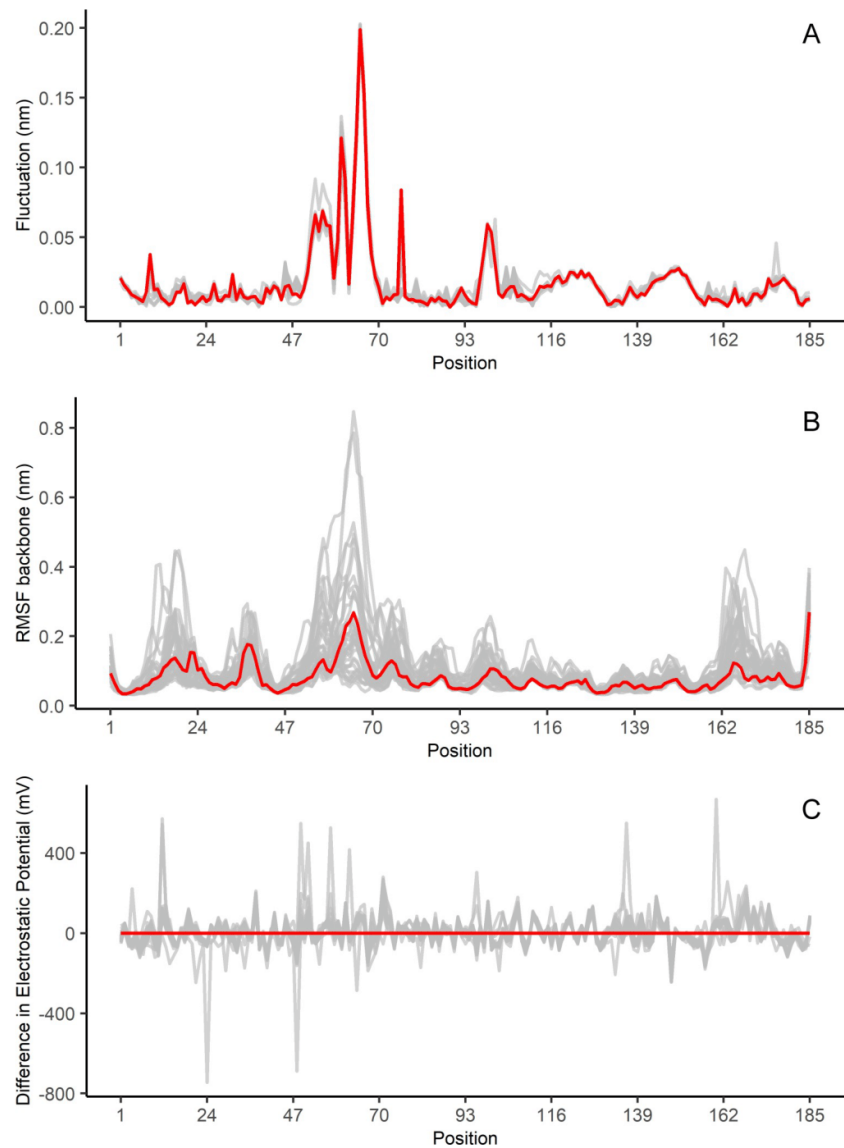
For the profile of stability per position, although sharing the same overall shape with the highest peak in the flap region going from His49 until His71, the fluctuations calculated with NMA (Fig 1A) showed a lower variability than the RMSF calculated with MD (Fig 1B). Concerning the DEPR, the profile also exhibits a high variability compared to the WT-PZase (Fig 1C), illustrating the effect in electrostatic potential introduced by missense mutations.

Comparing the different PZAses with a relative- $k_{cat}$  higher or equal than 50%, against the mutated PZAses with a relative- $k_{cat}$  lower than it, we found almost similar patterns of stability descriptors in both groups for NMA (S7A Fig and S7B Fig) and MD (S7C Fig and S7D Fig) approaches. In the case of physicochemical descriptors, a higher dispersion with outliers is observed for the group with relative- $k_{cat}$  less than 50 (S7E Fig and S7F Fig). After performing a two-sided Mann-Whitney-U non-parametric test for stability descriptors derived from NMA, we found statistically significant differences in the positions Leu27 (P-value = 0.022), Pro70 (P-value = 0.011), Tyr103 (P-value = 0.014), Val109 (P-value = 0.029), and Ser179 (P-value = 0.024). On the other hand, no statistically significant differences were found in any position for stability descriptors derived from MD simulations. Concerning DEPR, the positions Val45 (P-value = 0.026), Ala46 (P-value = 0.047), Asp49 (P-value = 0.032), His51 (P-value = 0.029), His57 (P-value = 0.029), Ser65 (P-value = 0.047), Arg121 (P-value = 0.024), Val131 (P-value = 0.047), and Asp166 (P-value = 0.016) showed a significant difference. Interestingly, three out of four residues of the MCS are included in that list. However, after correcting for multiple testing using the Bonferroni correction or the Benjamini-Hochberg procedure for an FDR of 10%, any significant position is found in any dataset.

### Log-linear models for kinetic parameters: $K_{cat}$ , $K_M$ , efficiency, and activity

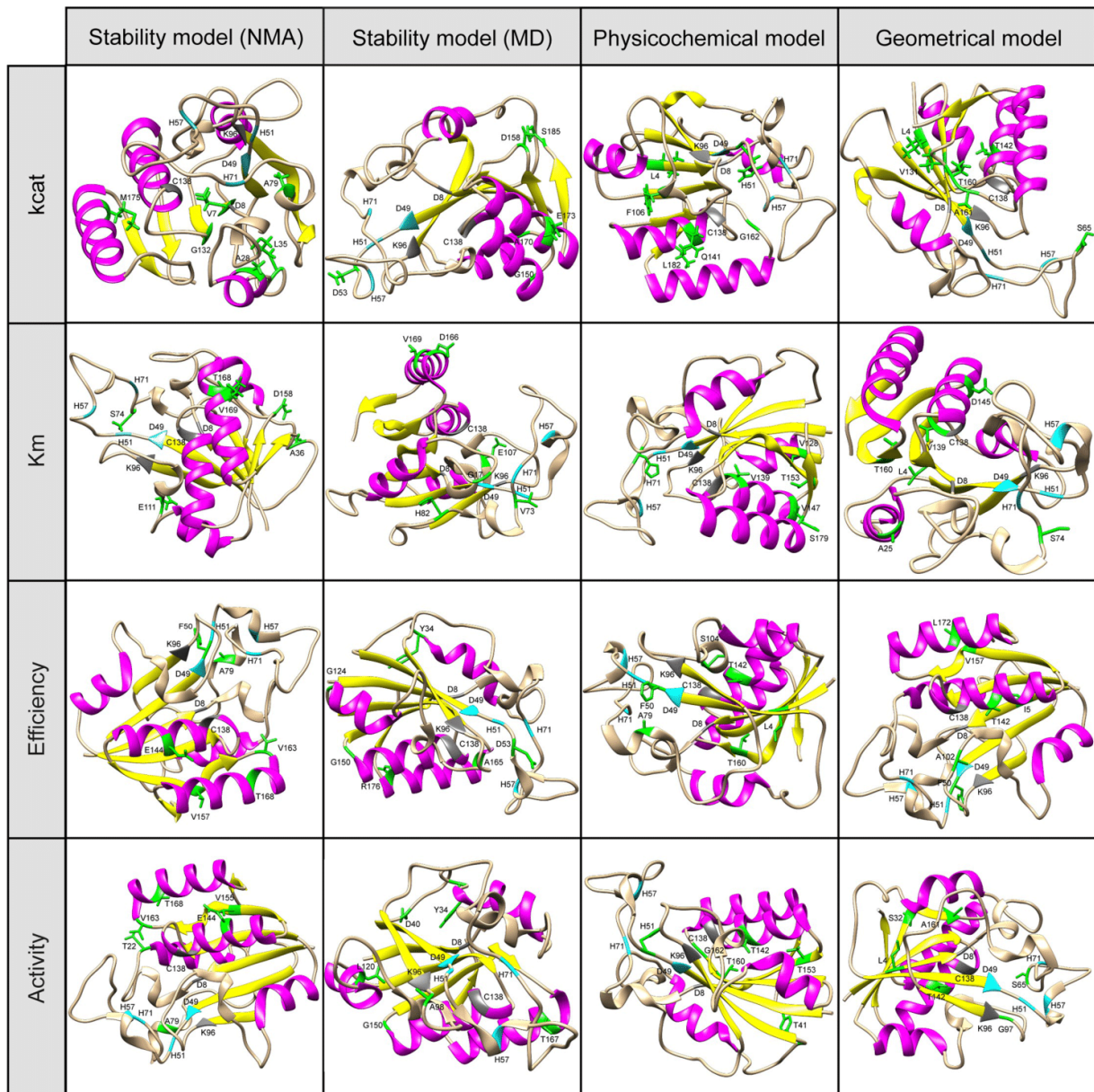
For each kinetic parameter, six log-linear models were constructed. Four of them correspond to individual linear models of six covariates of only physicochemical, geometrical, stability (NMA) or stability (MD) descriptors. The other two are weighted log-linear models built over the fitted values of the individual physicochemical, geometrical, and stability (NMA) or stability (MD) models.

The selected descriptors in individual models involve different positions along PZase (Fig 2 and Table 2). The Pearson correlation between descriptors in individual models is less than 0.8. We compared the distribution of the root-mean-square error (RMSE) of prediction calculated by 6-fold cross-validation for individual models against one of the random models of the



**Fig 1. Fluctuation, RMSF, and DEPR profiles for PZases obtained from normal mode analysis (NMA), molecular dynamics (MD), and MutantElec server.** (A) Fluctuation profiles. The red trace represents the WT profile. Low variability is observed for the fluctuation score along with all the positions. The overall shape of the profile is maintained along with the flap region spanning from His51 up to His71 being the most flexible part of PZase. (B) RMSF profiles. The red trace represents the WT profile. A higher variability than in the case of NMA fluctuations is observed. The flap region going from His51 up to His71 shows the maximum values for RMSF. Additionally, two other regions with high RMSF are observed near the N and C-terminal of the protein. (C) DEPR profiles. The red trace represents the WT profile. Differences greater than  $\pm 400$  mV are observed as extreme peaks, some of them belonging to the flap region of PZase.

<https://doi.org/10.1371/journal.pone.0235643.g001>



**Fig 2. Structures of WT-PZase highlighting the positions associated with the selected descriptors for individual models.** In green, residues from descriptor of the model; in gray, residues from the AS; in sky blue, residues from the MCS; in pink, residues from alpha-helix and in yellow, residues from beta-strands.

<https://doi.org/10.1371/journal.pone.0235643.g002>

same number and kind of descriptors. We found that individual models always have a mean RMSE lower than the random models, although the amount of the difference varies between different models (S8 Fig–S23 Fig).

Table 2. Selected descriptors for individual models of each kinetic parameter.

Selected descriptors	Stability model (NMA)	Kinetic parameters			
		$k_{cat}$	$K_M$	Efficiency	Activity
	Stability model (NMA)	Fluctuation Leu 35	Fluctuation Asp 158	<b>Fluctuation Thr 168</b>	<b>Fluctuation Thr 168</b>
		Fluctuation Met 175	Fluctuation Ala 36	<b>Fluctuation Ala 79</b>	<b>Fluctuation Ala 79</b>
		Fluctuation Gly 132	Fluctuation Glu 111	Fluctuation Val 157	<b>Fluctuation Glu 144</b>
		Fluctuation Val 7	Fluctuation Val 169	Fluctuation Phe 50	<b>Fluctuation Val 163</b>
		<b>Fluctuation Ala 79</b>	Fluctuation Ser 74	<b>Fluctuation Val 163</b>	Fluctuation Thr 22
		Fluctuation Ala 28	<b>Fluctuation Thr 168</b>	<b>Fluctuation Glu 144</b>	Fluctuation Val 155
	Stability model (MD-RMSF)	<b>RMSF Gly 150</b>	RMSF Gly 17	<b>RMSF Gly 150</b>	RMSF Asp 40
		<b>RMSF Gly 124</b>	RMSF Glu 107	<b>RMSF Gly 124</b>	<b>RMSF Gly 150</b>
		<b>RMSF Asp 53</b>	RMSF Val 169	RMSF Arg 176	RMSF Leu 120
		RMSF Glu 173	RMSF Asp 166	<b>RMSF Asp 53</b>	<b>RMSF Tyr 34</b>
		RMSF Ser 185	RMSF Val 73	<b>RMSF Tyr 34</b>	RMSF Thr 167
		RMSF Asp 158	RMSF His 82	RMSF Ala 165	RMSF Ala 98
	Physicochemical model	<b>DEPR Leu 4</b>	DEPR Val 139	<b>DEPR Leu 4</b>	<b>DEPR His 51</b>
		<b>DEPR His 51</b>	<b>DEPR Thr 153</b>	DEPR Phe 50	<b>DEPR Thr 160</b>
		DEPR Gln 141	DEPR Val 128	<b>DEPR Thr 160</b>	<b>DEPR Gly 162</b>
		DEPR Phe 106	DEPR Ser 179	DEPR Ala 79	<b>DEPR Thr 153</b>
		DEPR Leu 182	DEPR His 71	<b>DEPR Thr 142</b>	<b>DEPR Thr 142</b>
		<b>DEPR Gly 162</b>	DEPR Val 147	DEPR Ser 104	DEPR Tyr 41
	Geometrical model	<b>B_I (AS) Leu 4</b>	C_C138 Ser 74	<b>B_I (AS) Leu 4</b>	<b>B_I (AS) Thr 142</b>
		<b>B_I (AS) Thr 142</b>	<b>B_I (AS) Leu 4</b>	<b>B_I (AS) Thr 142</b>	P_C (MCS) Gly 97
		<b>C_C138 Ala 161</b>	B_P (MCS) Thr 160	<b>C_C138 Val 157</b>	<b>B_I (AS) Leu 4</b>
		<b>I_P (MCS) Ser 65</b>	C_K96 Ala 25	C_C138 Ala 102	<b>C_C138 Ala 161</b>
		C_D49 Val 131	C_D49 Asp 145	P_T (MCS) Leu 172	<b>I_P (MCS) Ser 65</b>
		I_C (MCS) Thr 160	B_I (AS) Val 139	I_P (MCS) Phe 50	<b>C_C138 Val 157</b>

The descriptors selected in more than one model are in bold. The meaning of geometrical descriptors is available at [S1 Appendix](#).

<https://doi.org/10.1371/journal.pone.0235643.t002>

An interesting result when comparing the alternative stability models is that RMSF descriptors generated by MD simulations always produce models with higher  $R^2$  and lower mean RMSE, probably due to the higher degree of variability between structures detected with this approach.

The weighted models have three covariates, representing the predictions of the stability, physicochemical and geometrical models for  $k_{cat}$  (S24 Fig and S25 Fig),  $K_M$  (S26 Fig and S27 Fig), efficiency (S28 Fig and S29 Fig) and activity (S30 Fig and S31 Fig). The coefficients for each covariate assign a weight to the prediction of the corresponding individual model. Overall, the six models for  $K_M$  have the lowest mean RMSE, always below 0.4. For  $k_{cat}$ , efficiency, and activity only the weighted models have a mean RMSE lower than 0.51. Regarding their individual models, the mean RMSE ranges from 0.906 to 1.37. A summary of the statistics of all the models for each kinetic parameter including  $R^2$ , adjusted  $R^2$ , p-value and mean RMSE is displayed in Table 3.

### Biological significance of the selected descriptors

Although the descriptors selected by individual models are not necessarily involved in the mechanism of action, we found that some positions associated with the selected descriptors are relatively close to the AS or the MCS. Thus, the distances between the  $C_\alpha$  of each residue in

Table 3. Summary statistics of models.

	$k_{cat}$				$K_M$			
	p-value	$R^2$	Adjusted $R^2$	<RMSE>	p-value	$R^2$	Adjusted $R^2$	<RMSE>
Weighted Model (NMA)	7.44E-16	0.897	0.887	0.5057	6.13E-14	0.864	0.851	0.1586
Weighted Model (MD)	8.70E-18	0.922	0.914	0.4299	3.20E-14	0.869	0.857	0.1562
Stability model (NMA)	7.76E-03	0.432	0.314	1.2514	6.93E-04	0.531	0.434	0.3441
Stability model (MD)	4.52E-05	0.619	0.54	1.0708	8.67E-06	0.662	0.593	0.2764
Physicochemical model	1.85E-08	0.784	0.739	0.9215	3.54E-08	0.773	0.727	0.247
Geometrical model	2.26E-11	0.866	0.838	1.3692	4.11E-08	0.771	0.724	0.265
	Efficiency				Activity			
	p-value	$R^2$	Adjusted $R^2$	<RMSE>	p-value	$R^2$	Adjusted $R^2$	<RMSE>
Weighted Model (NMA)	4.34E-17	0.92	0.912	0.4265	3.77E-16	0.901	0.892	0.4013
Weighted Model (MD)	6.80E-19	0.939	0.933	0.3707	4.45E-16	0.9	0.891	0.4001
Stability model (NMA)	4.51E-03	0.467	0.353	1.2189	2.05E-02	0.384	0.257	1.0487
Stability model (MD)	9.96E-06	0.671	0.601	0.9883	7.61E-05	0.603	0.521	0.9065
Physicochemical model	4.01E-10	0.845	0.812	1.1132	4.02E-08	0.771	0.724	0.91
Geometrical model	1.16E-12	0.899	0.877	1.2126	1.68E-08	0.785	0.741	1.0839

P-value refers to significance;  $R^2$  is the coefficient of determination and <RMSE> is the mean of Root-Mean-Square Error.

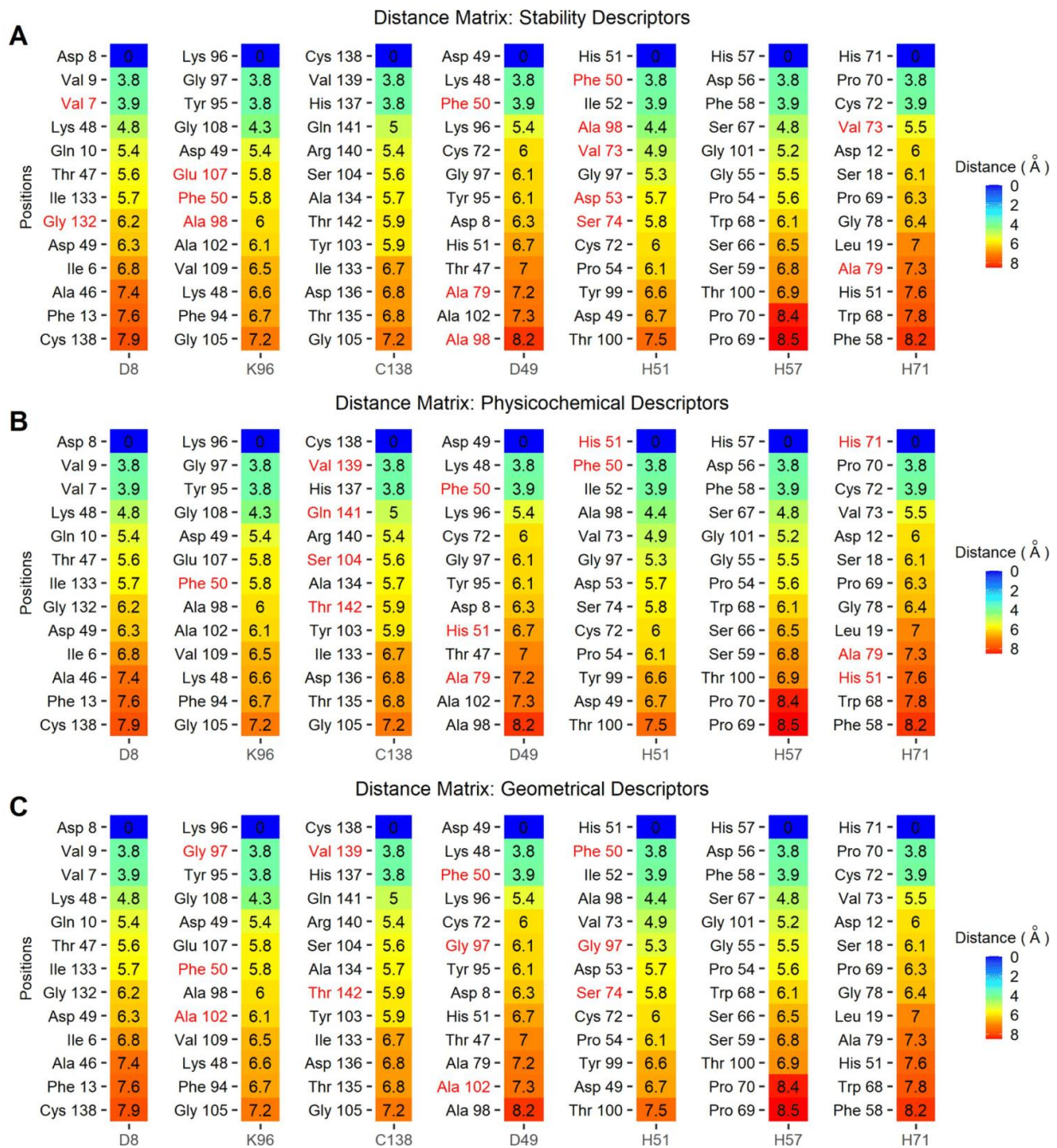
<https://doi.org/10.1371/journal.pone.0235643.t003>

the WT-PZase and the  $C_{\alpha}$  of the residues of the AS and MCS were calculated and sorted to highlight the closest positions to these points (Fig 3 and S2 Table). To select the closest positions a cutoff of 7 Å was arbitrarily chosen. The residues with positions included in the selected descriptors close to at least one of the residues of the AS or MCS are Phe50, Ser74, Gly97, Ala102, Val139, and Thr142 for the geometrical models, Phe50, His51, His71, Ser104, Val139, Gln141, and Thr142 for the physicochemical models, and Val7, Phe50, Asp53, Val73, Ser74, Ala98, Glu107, and Gly132 and for the stability models (NMA and MD). These positions reflect the importance of the stability, electrostatic potential, and geometrical distances near the AS/MCS, while the rest of positions the effect of long-range interactions in the kinetics parameters of PZase.

## Discussion

In this study, we developed a method to predict PZase kinetic parameters ( $k_{cat}$ ,  $K_M$ , efficiency, and activity) based on recognizing a mutation in the *pncA* coding gene sequence. Stability, physicochemical, and geometrical descriptors got from the modeled 3D structure were used to build multiple log-linear models to predict the kinetic parameters. Our weighted models can predict the experimentally measured enzymatic kinetic parameters with high accuracy, with adjusted  $R^2$  values ranging from 85.1% to 93.3%. These models confirmed to be robust after a 6-fold cross-validation analysis, with the models for activity and  $K_M$  having the lowest mean RMSE.

The kinetic parameters studied belong to two groups. A group related to the catalytic activity ( $k_{cat}$ , efficiency, and activity) and a group related to substrate affinity ( $K_M$ ). Even though efficiency depends inversely on  $K_M$ , it showed to be highly correlated with  $k_{cat}$  and activity. While a low value for the group related to catalytic activity has a deleterious effect in the enzymatic reaction, the opposite is true for the group related to substrate affinity, where a high value of  $K_M$  implies that a higher concentration of substrate is needed to have the enzyme working at its maximum velocity. In our data, the distribution of  $K_M$  showed that most mutants have similar values to the WT-PZase, while only six mutants (D49N, D136G, V180F,



**Fig 3. Distance matrix of WT-PZase residues position against AS and MCS.** The distance between pairs of  $C_{\alpha}$  is sorted ascendingly. The matrix shows the 13 positions closest to the residues of the AS and MCS. The residues whose position is included in the selected descriptors are depicted in red for (A) stability models (NMA and MD), (B) physicochemical models, and (C) geometrical models.

<https://doi.org/10.1371/journal.pone.0235643.g003>

H71Y, T160K, and V139A) have values of  $K_M$  2-fold greater than the WT. For  $k_{cat}$  efficiency, and activity, several mutants have very low values, showing that missense mutations affect catalytic activity rather than substrate affinity most of the time.

Regarding the physicochemical descriptors, the cavity formed by the AS and the MCS of PZase is rich in residues with a negative charge at neutral pH. At physiological pH, POA and PZA are negatively charged (pKa 2.9 and 0.5 respectively [72]). Based on this evidence, we hypothesize that an increase in the DEPR (compared to the WT-PZase) at positions close to this cavity will increase the affinity for PZA of the enzyme but decrease the release rate of the POA. This is important to consider, since PZA is a prodrug and does not act as an inhibitor, but needs to be hydrolyzed into POA and then released from the PZase to perform its lethal effect.

Theoretically, increasing or decreasing DEPR may eventually cause enzymatic dysfunction. Interestingly, variation in the DEPR of positions His51 and His71 from MCS are negatively correlated with  $k_{cat}$  and  $K_M$ , respectively, showing that a higher DEPR will reduce the catalytic activity (slower release) and the dissociation constant (stronger interaction, lower amounts of substrate needed to saturate the enzyme). Also, for DEPR, three positions belonging to the MCS (Asp49, His51, and His57) showed a statistically significant difference, suggesting that the difference in electrostatic potentials could be informative for classifying functional and non-functional PZases. To confirm this hypothesis, further experimental studies are required to evaluate a larger number of mutants.

Our stability data from NMA and MD approaches shows that the fluctuation profile is overall conserved among PZases, with the flap region from His51 to His71 being the most flexible part of the enzyme, as previously reported [45–49,73]. This loop is in the binding's lid cavity and contributes to the ability of PZA to enter the PZase active site cavity [13]. At the prediction level, stability predicted by MD simulations performed better than that predicted by NMA. This could be because of the different degrees of variation detected by NMA and MD. While NMA detects almost no variation in the fluctuation profiles of PZases (S4 Fig), MD allows a better resolution of the effect of punctual mutations in the stability of PZase. However, in terms of prediction time, NMA based stability and weighted models will provide much faster estimates than MD based models [51].

After dichotomizing the PZases in two groups using a value of relative- $k_{cat}$  of 50% as a cut-off (high relative- $k_{cat}$  group and low relative- $k_{cat}$  group), we did not find statistically significant differences in the RMSF values, but for the fluctuation scores, only one position from the flap region (Pro70) was significant. Although previous studies showed that PZase function decreases with an increment in the RMSF of the flap region, [42–46,74], we did not find statistical evidence for that. This apparent discrepancy may occur because other studies only compared the fluctuation profiles visually with few data or performing no statistical test.

A previous study from our group reported a predictive model based on 16 geometrical descriptors that predicted PZase kinetic parameters, explaining approximately 87% of the variability in  $k_{cat}$  [38]. Compared to our previous study and based on the recent availability of the WT-PZase crystal structure [13], we considered Lys96 as part of the AS, and His57 as part of the MCS. Therefore, this study is a considerable improvement in our previous work, plus it includes stability and physicochemical aspects, besides the geometric descriptors. In contrast to previous studies based on a small number of mutants [42–46,74–76], our study includes a deeply statistical analysis based on 500 nanoseconds MD of every mutant PZase. Therefore, discrepancies are expected to occur.

PZase is a metalloenzyme that *in-vitro* could be activated by the coordination of several ions in the MCS [77,78]. Co, Cd, and Mn are the most important metal cofactors *in vitro*, although they may not have a significant effect *in vivo* given its low abundance in the

intracellular environment. *In vivo*, previous studies found Fe [78] and Zn [77] to be coordinated with the PZase in MTB. Therefore, it is likely that several metals may coordinate PZase *in vivo*, with different abundances. Because of the lack of knowledge of this exact distribution of coordinated metals, plus the high computational demand for including the several 3D structural models for each metal coordinated, we decided as a first approach to only consider the apo-PZase structure, in our calculations. We believe that the correct selection of metal coordinated PZase and quantum mechanics calculations [79,80], may improve the prediction of the enzymatic function. For this, further studies are required.

There are several limitations during the development of a predictive model as the one described here. The experimental measurements of the kinetic parameters are highly variable between different batches; however, we controlled this by normalizing the parameters against the corresponding to the WT-PZase. Also, available information about mutated PZases and the measurement of its kinetic parameters is scarce compared with information related to the strain susceptibility. Also, we have focused our models on predicting enzymatic parameters of point mutations in PZase without considering insertions or deletions, because the effects of this kind of mutations are not reliably predicted.

Based on this data, several studies pretended to dichotomously predict PZA susceptibility (resistant or susceptible) only from mutations in PZase detected by *pncA* sequencing [81–83]. These approaches may be biased, with a risk of missing the correct physical/biological implications of PZase mutations because PZA resistance could also be attributed to other factors besides PZase activity itself, like differential PZase expression levels [84] or dysfunction in other targets like *panD* [85], and many others still unknown [86]. This problem is explicit when a strain has PZase mutations that do not affect the PZase enzymatic function, but mutations in other critical genes [14,87,88] that generate PZA resistance in the bacteria. These cases would erroneously attribute PZase mutations that do not affect the enzymatic function, the apparent effect of causing PZA resistance. This approach could be improved using whole-genome sequencing [89–92], but the expression and environmental factors are still present.

For these reasons, and considering that resistance to PZA can be attributed to multiple mechanisms, it is important to first predict PZase enzymatic function from the *pncA* sequence. The only certain interpretation is to infer resistance to PZA when the *pncA* mutation causes a significant loss of enzymatic function. However, those mutations that do not significantly affect function, do not necessarily predict that the bacterium is susceptible, as mutations in genes associated with other resistance mechanisms may be present. In these cases, it is important to analyze mutations in other genes associated with alternative mechanisms of resistance to PZA.

In conclusion, the present work we show the construction of log-linear models based on geometrical, physicochemical, and stability descriptors derived from the modeled structure of the *M. tuberculosis* PZase. This is useful to predict functional kinetic parameters of PZase related to PZA resistance, based on the high certainty of PZase dysfunction, from only the *pncA* gene sequence. This can be an important tool to contribute to efforts to detect early resistance to PZA.

## Supporting information

### S1 Fig. Ribbon representation of the structure of the *M. tuberculosis* PZase protein.

Highlighted in pink, residues from alpha helixes; in yellow, residues from beta-strands and in orange, residues from the flap region.

(TIF)

**S2 Fig. Tridimensional representations of geometrical reference points.** (A) Points related to the active site (AS) and (B) points related to the metal coordination site (MCS).  $V_r$  represents the resultant vector of each residue.

(TIF)

**S3 Fig. DEPR profiles for WT and thirty-five mutated PZAses.** Profiles of wild-type and mutated PZAses sorted by relative- $k_{cat}$ , each plot represent the difference in electrostatic potential respect to the wild-type for a given position.

(TIF)

**S4 Fig. NMA fluctuation profiles for WT and thirty-five mutated PZAses.** Profiles of wild-type and mutated PZAses sorted by relative- $k_{cat}$ .

(TIF)

**S5 Fig. MD RMSD profiles for WT and thirty-five mutated PZAses.** Profiles of wild-type and mutated PZAses sorted by relative- $k_{cat}$ , trajectories of 500 ns molecular dynamics of the entire protein were used as input.

(TIF)

**S6 Fig. MD RMSF profiles for WT and thirty-five mutated PZAses.** Profiles of wild-type and mutated PZAses sorted by relative- $k_{cat}$ , trajectories of the last 100ns of a 500ns molecular dynamics of the protein backbone were used as input.

(TIF)

**S7 Fig. Per-position Boxplots for stability and physicochemical descriptors.** In blue, PZAses with a  $k_{cat}$  greater or equal than 50. In red, PZAses with a  $k_{cat}$  lower than 50. (A, B) Fluctuations from NMA analysis. (C, D) RMSFs of the last 100ns from MD analysis (E, F) DEPRs from the MutantElec server.

(TIF)

**S8 Fig. Stability model (NMA) for  $k_{cat}$ .** (A) Table with estimated coefficients and statistics for the selected stability descriptors (fluctuations). (B) Comparison of statistics ( $R^2$ , Adjusted  $R^2$ , P-value, and RMSE) between the stability model and a random stability model. (C) Fitted values and experimental values for mean  $\log_{10}$  (relative- $k_{cat}$ ). (D) Heatmap showing the correlation coefficient between the selected descriptors. (E) Confidence intervals for the coefficients of each stability descriptor. (F) Distribution of RMSEs calculated by 6-fold cross-validation for the stability model (red) and a random model (blue).

(TIF)

**S9 Fig. Stability model (MD) for  $k_{cat}$ .** (A) Table with estimated coefficients and statistics for the selected stability descriptors (RMSFs). (B) Comparison of statistics ( $R^2$ , Adjusted  $R^2$ , P-value, and RMSE) between the stability model and a random stability model. (C) Fitted values and experimental values for mean  $\log_{10}$  (relative- $k_{cat}$ ). (D) Heatmap showing the correlation coefficient between the selected descriptors. (E) Confidence intervals for the coefficients of each stability descriptor. (F) Distribution of RMSEs calculated by 6-fold cross-validation for the stability model (red) and a random model (blue).

(TIF)

**S10 Fig. Physicochemical model for  $k_{cat}$ .** (A) Table with estimated coefficients and statistics for the selected physicochemical descriptors (DEPRs) (B) Comparison of statistics ( $R^2$ , Adjusted  $R^2$ , P-value, and RMSE) between the physicochemical model and a random physicochemical model. (C) Fitted values and experimental values for mean  $\log_{10}$  (relative- $k_{cat}$ ). (D) Heatmap showing the correlation coefficient between the selected descriptors. (E) Confidence

intervals for the coefficients of each physicochemical descriptor. (F) Distribution of RMSEs calculated by 6-fold cross-validation for the physicochemical model (red) and a random model (blue).

(TIF)

**S11 Fig. Geometrical model for  $k_{cat}$ .** (A) Table with estimated coefficients and statistics for the selected geometrical descriptors (B) Comparison of statistics ( $R^2$ , Adjusted  $R^2$ , P-value, and RMSE) between the geometrical model and a random geometrical model. (C) Fitted values and experimental values for mean  $\log_{10}$  (relative- $k_{cat}$ ). (D) Heatmap showing the correlation coefficient between the selected descriptors. (E) Confidence intervals for the coefficients of each geometrical descriptor. (F) Distribution of RMSEs calculated by 6-fold cross-validation for the geometrical model (red) and a random model (blue).

(TIF)

**S12 Fig. Stability model (NMA) for  $K_M$ .** (A) Table with estimated coefficients and statistics for the selected stability descriptors (fluctuations), (B) Comparison of statistics ( $R^2$ , Adjusted  $R^2$ , P-value, and RMSE) between the stability model and a random stability model. (C) Fitted values and experimental values for mean  $\log_{10}$  (relative- $K_M$ ). (D) Heatmap showing the correlation coefficient between the selected descriptors. (E) Confidence intervals for the coefficients of each stability descriptor. (F) Distribution of RMSEs calculated by 6-fold cross-validation for the stability model (red) and a random model (blue).

(TIF)

**S13 Fig. Stability model (MD) for  $K_M$ .** (A) Table with estimated coefficients and statistics for the selected stability descriptors (RMSFs). (B) Comparison of statistics ( $R^2$ , Adjusted  $R^2$ , P-value, and RMSE) between the stability model and a random stability model. (C) Fitted values and experimental values for mean  $\log_{10}$  (relative- $K_M$ ). (D) Heatmap showing the correlation coefficient between the selected descriptors. (E) Confidence intervals for the coefficients of each stability descriptor. (F) Distribution of RMSEs calculated by 6-fold cross-validation for the stability model (red) and a random model (blue).

(TIF)

**S14 Fig. Physicochemical model for  $K_M$ .** (A) Table with estimated coefficients and statistics for the selected physicochemical descriptors (DEPRs). (B) Comparison of statistics ( $R^2$ , Adjusted  $R^2$ , P-value, and RMSE) between the physicochemical model and a random physicochemical model. (C) Fitted values and experimental values for mean  $\log_{10}$  (relative- $K_M$ ). (D) Heatmap showing the correlation coefficient between the selected descriptors. (E) Confidence intervals for the coefficients of each physicochemical descriptor. (F) Distribution of RMSEs calculated by 6-fold cross-validation for the physicochemical model (red) and a random model (blue).

(TIF)

**S15 Fig. Geometrical model for  $K_M$ .** (A) Table with estimated coefficients and statistics for the selected geometrical descriptors. (B) Comparison of statistics ( $R^2$ , Adjusted  $R^2$ , P-value, and RMSE) between the geometrical model and a random geometrical model. (C) Fitted values and experimental values for mean  $\log_{10}$  (relative- $K_M$ ). (D) Heatmap showing the correlation coefficient between the selected descriptors. (E) Confidence intervals for the coefficients of each geometrical descriptor. (F) Distribution of RMSEs calculated by 6-fold cross-validation for the geometrical model (red) and a random model (blue).

(TIF)

**S16 Fig. Stability model (NMA) for efficiency.** (A) Table with estimated coefficients and statistics for the selected stability descriptors (fluctuations). (B) Comparison of statistics ( $R^2$ , Adjusted  $R^2$ , P-value, and RMSE) between the stability model and a random stability model. (C) Fitted values and experimental values for mean  $\log_{10}$  (relative efficiency). (D) Heatmap showing the correlation coefficient between the selected descriptors. (E) Confidence intervals for the coefficients of each stability descriptor. (F) Distribution of RMSEs calculated by 6-fold cross-validation for the stability model (red) and a random model (blue). (TIF)

**S17 Fig. Stability model (MD) for efficiency.** (A) Table with estimated coefficients and statistics for the selected stability descriptors (RMSFs). (B) Comparison of statistics ( $R^2$ , Adjusted  $R^2$ , P-value, and RMSE) between the stability model and a random stability model. (C) Fitted values and experimental values for mean  $\log_{10}$  (relative efficiency). (D) Heatmap showing the correlation coefficient between the selected descriptors. (E) Confidence intervals for the coefficients of each stability descriptor. (F) Distribution of RMSEs calculated by 6-fold cross-validation for the stability model (red) and a random model (blue). (TIF)

**S18 Fig. Physicochemical model for efficiency.** (A) Table with estimated coefficients and statistics for the selected physicochemical descriptors (DEPRs). (B) Comparison of statistics ( $R^2$ , Adjusted  $R^2$ , P-value, and RMSE) between the physicochemical model and a random physicochemical model. (C) Fitted values and experimental values for mean  $\log_{10}$  (relative efficiency). (D) Heatmap showing the correlation coefficient between the selected descriptors. (E) Confidence intervals for the coefficients of each physicochemical descriptor. (F) Distribution of RMSEs calculated by 6-fold cross-validation for the physicochemical model (red) and a random model (blue). (TIF)

**S19 Fig. Geometrical model for efficiency.** (A) Table with estimated coefficients and statistics for the selected geometrical descriptors. (B) Comparison of statistics ( $R^2$ , Adjusted  $R^2$ , P-value, and RMSE) between the geometrical model and a random geometrical model. (C) Fitted values and experimental values for mean  $\log_{10}$  (relative efficiency). (D) Heatmap showing the correlation coefficient between the selected descriptors. (E) Confidence intervals for the coefficients of each geometrical descriptor. (F) Distribution of RMSEs calculated by 6-fold cross-validation for the geometrical model (red) and a random model (blue). (TIF)

**S20 Fig. Stability model (NMA) for activity.** (A) Table with estimated coefficients and statistics for the selected stability descriptors (fluctuations). (B) Comparison of statistics ( $R^2$ , Adjusted  $R^2$ , P-value, and RMSE) between the stability model and a random stability model. (C) Fitted values and experimental values for mean  $\log_{10}$  (relative activity). (D) Heatmap showing the correlation coefficient between the selected descriptors. (E) Confidence intervals for the coefficients of each stability descriptor. (F) Distribution of RMSEs calculated by 6-fold cross-validation for the stability model (red) and a random model (blue). (TIF)

**S21 Fig. Stability model (MD) for activity.** (A) Table with estimated coefficients and statistics for the selected stability descriptors (RMSFs). (B) Comparison of statistics ( $R^2$ , Adjusted  $R^2$ , P-value, and RMSE) between the stability model and a random stability model. (C) Fitted values and experimental values for mean  $\log_{10}$  (relative activity). (D) Heatmap showing the correlation coefficient between the selected descriptors. (E) Confidence intervals for the coefficients

of each stability descriptor. (F) Distribution of RMSEs calculated by 6-fold cross-validation for the stability model (red) and a random model (blue).

(TIF)

**S22 Fig. Physicochemical model for activity.** (A) Table with estimated coefficients and statistics for the selected physicochemical descriptors (DEPRs). (B) Comparison of statistics ( $R^2$ , Adjusted  $R^2$ , P-value, and RMSE) between the physicochemical model and a random physicochemical model. (C) Fitted values and experimental values for mean  $\log_{10}$  (relative activity). (D) Heatmap showing the correlation coefficient between the selected descriptors. (E) Confidence intervals for the coefficients of each physicochemical descriptor. (F) Distribution of RMSEs calculated by 6-fold cross-validation for the physicochemical model (red) and a random model (blue).

(TIF)

**S23 Fig. Geometrical model for activity.** (A) Table with estimated coefficients and statistics for the selected geometrical descriptors. (B) Comparison of statistics ( $R^2$ , Adjusted  $R^2$ , P-value, and RMSE) between the geometrical model and a random geometrical model. (C) Fitted values and experimental values for mean  $\log_{10}$  (relative activity). (D) Heatmap showing the correlation coefficient between the selected descriptors. (E) Confidence intervals for the coefficients of each geometrical descriptor. (F) Distribution of RMSEs calculated by 6-fold cross-validation for the geometrical model (red) and a random model (blue).

(TIF)

**S24 Fig. Weighted model (NMA) for  $k_{cat}$ .** (A) Table with estimated coefficients and statistics for the individual predictions of stability (NMA), physicochemical and geometrical models. (B) Comparison among the individual models and the weighted model (NMA) for  $k_{cat}$ . (C) Fitted values and experimental values for mean  $\log_{10}$  (relative- $k_{cat}$ ) for the weighted model (NMA). (D) Distribution of RMSE calculated by 6-fold cross-validation for the weighted and individual models.

(TIF)

**S25 Fig. Weighted model (MD) for  $k_{cat}$ .** (A) Table with estimated coefficients and statistics for the individual predictions of stability (MD), physicochemical and geometrical models. (B) Comparison among the individual models and the weighted model (MD) for  $k_{cat}$ . (C) Fitted values and experimental values for mean  $\log_{10}$  (relative- $k_{cat}$ ) for the weighted model (MD). (D) Distribution of RMSE calculated by 6-fold cross-validation for the weighted and individual models.

(TIF)

**S26 Fig. Weighted model (NMA) for  $K_M$ .** (A) Table with estimated coefficients and statistics for the individual predictions of stability (NMA), physicochemical and geometrical models. (B) Comparison among the individual models and the weighted model (NMA) for  $K_M$ . (C) Fitted values and experimental values for mean  $\log_{10}$  (relative- $K_M$ ) for the weighted model (NMA). (D) Distribution of RMSE calculated by 6-fold cross-validation for the weighted and individual models.

(TIF)

**S27 Fig. Weighted model (MD) for  $K_M$ .** (A) Table with estimated coefficients and statistics for the individual predictions of stability (MD), physicochemical and geometrical models. (B) Comparison among the individual models and the weighted model (MD) for  $K_M$ . (C) Fitted values and experimental values for mean  $\log_{10}$  (relative- $K_M$ ) for the weighted model (MD). (D) Distribution of RMSE calculated by 6-fold cross-validation for the weighted and individual

models.  
(TIF)

**S28 Fig. Weighted model (NMA) for efficiency.** (A) Table with estimated coefficients and statistics for the individual predictions of stability (NMA), physicochemical and geometrical models. (B) Comparison among the individual models and the weighted model (NMA) for efficiency. (C) Fitted values and experimental values for mean  $\log_{10}$  (relative efficiency) for the weighted model (NMA). (D) Distribution of RMSE calculated by 6-fold cross-validation for the weighted and individual models.  
(TIF)

**S29 Fig. Weighted model (MD) for efficiency.** (A) Table with estimated coefficients and statistics for the individual predictions of stability (MD), physicochemical and geometrical models. (B) Comparison among the individual models and the weighted model (MD) for efficiency. (C) Fitted values and experimental values for mean  $\log_{10}$  (relative efficiency) for the weighted model (MD). (D) Distribution of RMSE calculated by 6-fold cross-validation for the weighted and individual models.  
(TIF)

**S30 Fig. Weighted model (NMA) for activity.** (A) Table with estimated coefficients and statistics for the individual predictions of stability (NMA), physicochemical and geometrical models. (B) Comparison among the individual models and the weighted model (NMA) for activity. (C) Fitted values and experimental values for mean  $\log_{10}$  (relative activity) for the weighted model (NMA). (D) Distribution of RMSE calculated by 6-fold cross-validation for the weighted and individual models.  
(TIF)

**S31 Fig. Weighted model (MD) for activity.** (A) Table with estimated coefficients and statistics for the individual predictions of stability (MD), physicochemical and geometrical models. (B) Comparison among the individual models and the weighted model (MD) for activity. (C) Fitted values and experimental values for mean  $\log_{10}$  (relative activity) for the weighted model (MD). (D) Distribution of RMSE calculated by 6-fold cross-validation for the weighted and individual models.  
(TIF)

**S1 Dataset. Dataset of geometrical descriptors per residue of WT-PZAse and 35 mutants.** In gray, mutants from the second batch.  
(XLSX)

**S2 Dataset. Dataset of physicochemical descriptors per residue of WT-PZAse and 35 mutants.** In gray, mutants from the second batch.  
(XLSX)

**S3 Dataset. Dataset of fluctuations per residue from NMA analysis of WT-PZAse and 35 mutants.** In gray, mutants from the second batch.  
(XLSX)

**S4 Dataset. Dataset of fluctuations per residue from MD analysis of WT-PZAse and 35 mutants.** In gray, mutants from the second batch.  
(XLSX)

**S1 Table. Absolute and relative standard errors for kinetic parameters of two batches of mutated PZAses from *M. tuberculosis*.**

(XLSX)

**S2 Table. The distance matrix for selected residues for each model to the residues of AS and MCS.** The descriptors that were selected in more than one model are in bold.

(XLSX)

**S1 Appendix. List of geometrical descriptors used in geometrical modeling.** A generalized meaning of each of them.

(XLSX)

## Acknowledgments

We are grateful to Basilio Cieza for the facilitation and initial instruction in the use of the vibrational normal mode analysis scripts of Dr. Amzel's group.

## Author Contributions

**Conceptualization:** Robert H. Gilman, Patricia Sheen, Mirko Zimic.

**Data curation:** Rydberg Roman Supo-Escalante, Aldhair Médico.

**Formal analysis:** Rydberg Roman Supo-Escalante, Aldhair Médico, Eduardo Gushiken, Mirko Zimic.

**Funding acquisition:** Patricia Sheen, Mirko Zimic.

**Investigation:** Rydberg Roman Supo-Escalante, Aldhair Médico, Eduardo Gushiken, Mirko Zimic.

**Methodology:** Rydberg Roman Supo-Escalante, Aldhair Médico, Eduardo Gushiken, Gustavo E. Olivos-Ramírez, Yaneth Quispe, Fiorella Torres, Melissa Zamudio, Ricardo Antiparra, Patricia Sheen, Mirko Zimic.

**Project administration:** Mirko Zimic.

**Resources:** Patricia Sheen, Mirko Zimic.

**Software:** Rydberg Roman Supo-Escalante, Aldhair Médico, Eduardo Gushiken, Gustavo E. Olivos-Ramírez.

**Supervision:** Eduardo Gushiken, L. Mario Amzel, Patricia Sheen, Mirko Zimic.

**Visualization:** Rydberg Roman Supo-Escalante, Aldhair Médico.

**Writing – original draft:** Rydberg Roman Supo-Escalante, Aldhair Médico, Patricia Sheen, Mirko Zimic.

**Writing – review & editing:** Rydberg Roman Supo-Escalante, Aldhair Médico, Eduardo Gushiken, Gustavo E. Olivos-Ramírez, L. Mario Amzel, Robert H. Gilman, Patricia Sheen, Mirko Zimic.

## References

1. WHO—World Health Organization. World health statistics 2018: monitoring health for the SDGs, sustainable development goals. 2018.