

RESEARCH ARTICLE

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

Rydberg Roman Supo-Escalante¹✉, Aldhair Médico¹✉, Eduardo Gushiken¹✉, Gustavo E. Olivos-Ramírez¹✉, Yaneth Quispe¹✉, Fiorella Torres¹✉, Melissa Zamudio¹✉, Ricardo Antiparra¹✉, L. Mario Amzel²✉, Robert H. Gilman³✉, Patricia Sheen¹✉, Mirko Zimic¹*✉

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



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 to 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/ Fe² binding cavity [13], and alterations of its stability have been reported in mutated PZases 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 μ 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 μ L of 20% Ferrous ammonium sulfate was added, followed immediately by adding 445 μ 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 μ 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 Å 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 Å. 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 7th up to the 46th 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⁺ and Cl⁻ 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⁻¹. 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 (ϵ_{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 PZases

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 α 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 (Vr) of each amino acid regarding the C α 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 σ_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} (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
	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
	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	0.000	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
	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	1099.194	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	2050.140	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	
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	
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

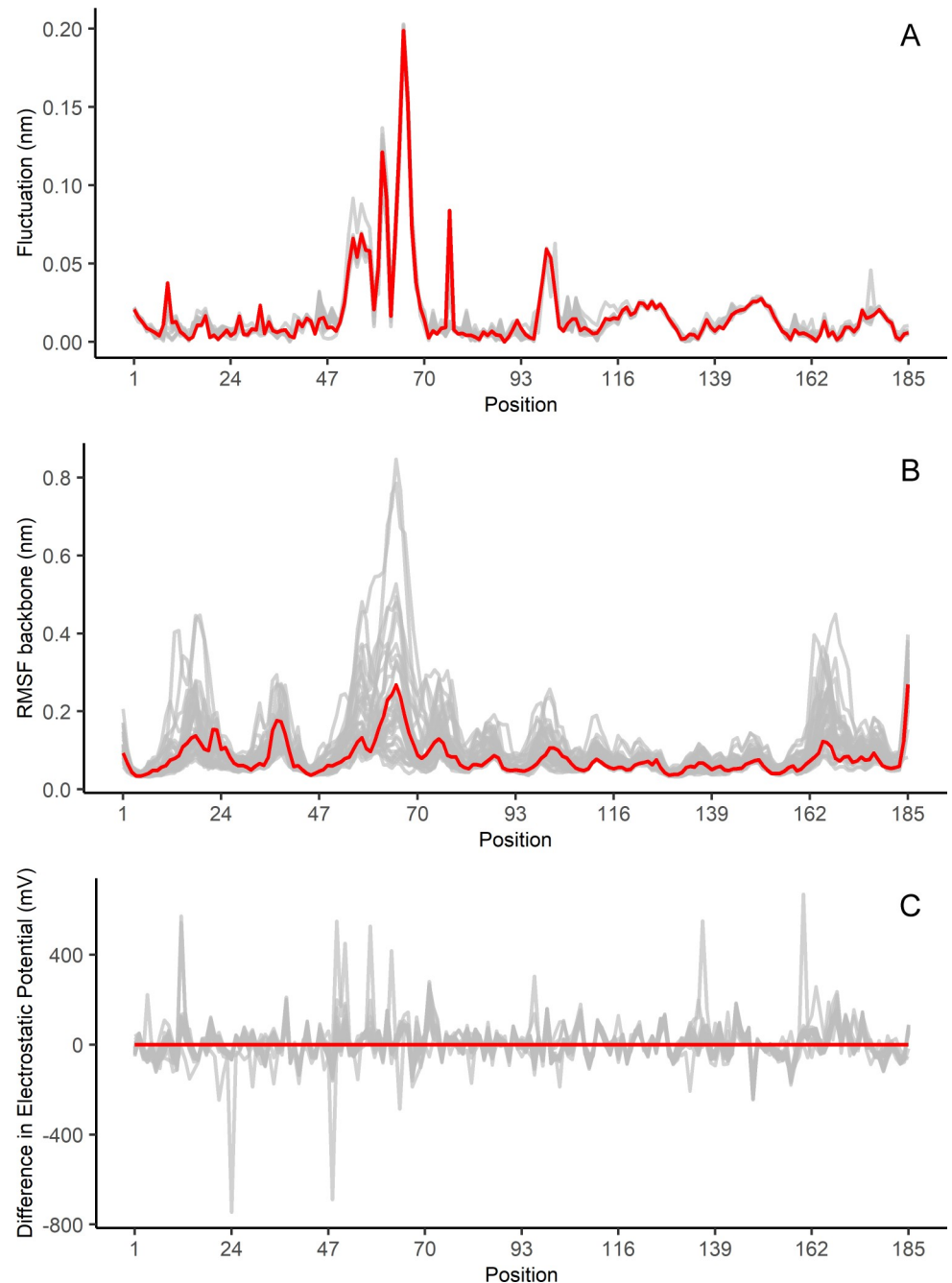


Fig 1. Fluctuation, RMSF, and DEPR profiles for PZAses obtained from normal mode analysis (NMA), molecular dynamics (MD), and MutantElec server. (A) Fluctuation profiles. The red trace represents the WT profile. Low variability is observed for the fluctuation score along with all the positions. The overall 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 ± 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>

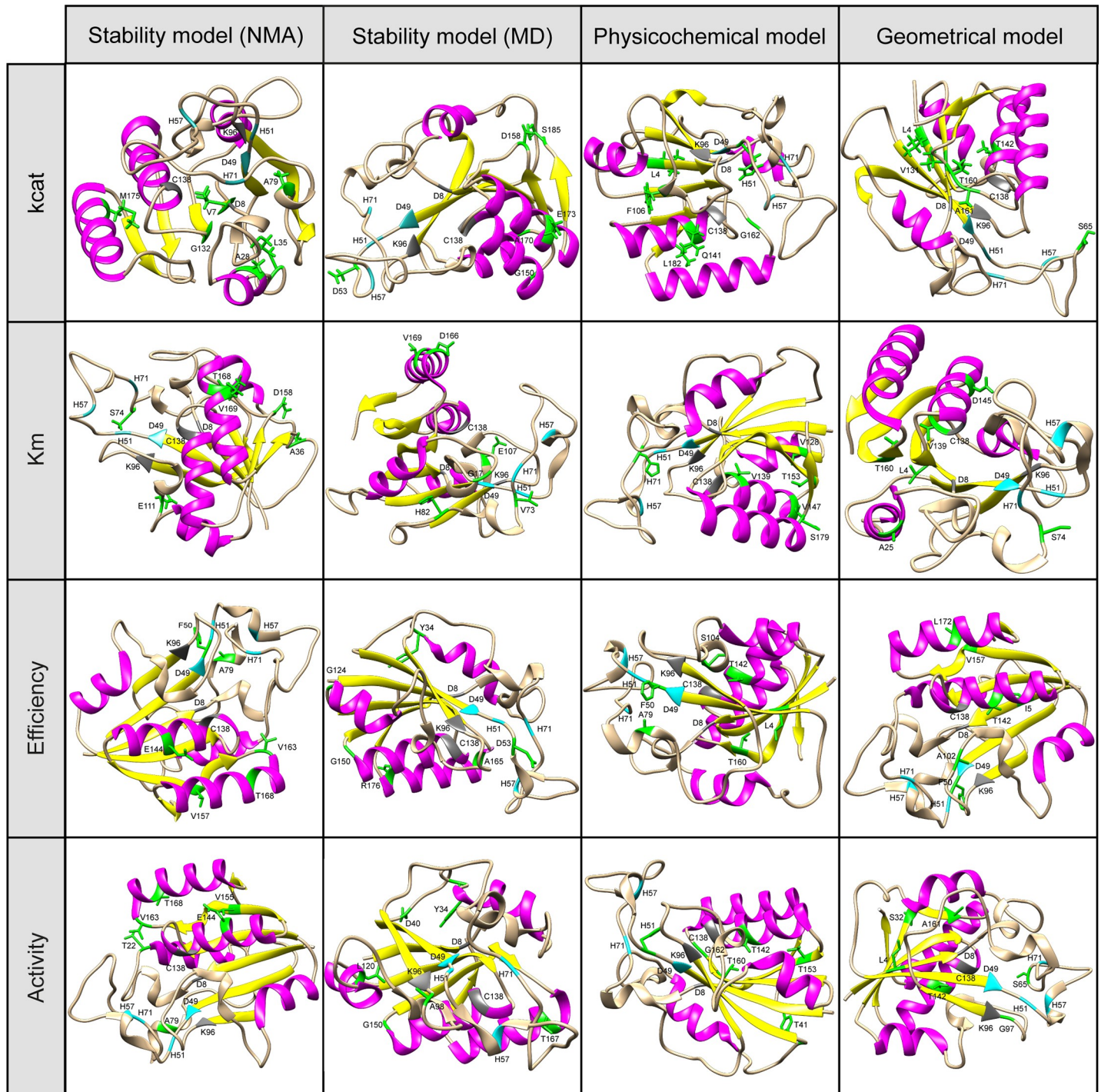


Fig 2. Structures of WT-PZase highlighting the positions associated with the selected descriptors for individual models. In green, residues from each descriptor of the model; in gray, residues from the AS; in sky blue, residues from the MCS; in pink, residues from alpha-helix and in yellow, residues 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	Model	Kinetic parameters			
		k_{cat}	K_M	Efficiency	Activity
Selected descriptors	Stability model (NMA)	Fluctuation Leu 35	Fluctuation Asp 158	Fluctuation Thr 168	Fluctuation Thr 168
		Fluctuation Met 175	Fluctuation Ala 36	Fluctuation Ala 79	Fluctuation Ala 79
		Fluctuation Gly 132	Fluctuation Glu 111	Fluctuation Val 157	Fluctuation Glu 144
		Fluctuation Val 7	Fluctuation Val 169	Fluctuation Phe 50	Fluctuation Val 163
		Fluctuation Ala 79	Fluctuation Ser 74	Fluctuation Val 163	Fluctuation Thr 22
	Stability model (MD-RMSF)	Fluctuation Ala 28	Fluctuation Thr 168	Fluctuation Glu 144	Fluctuation Val 155
		RMSF Gly 150	RMSF Gly 17	RMSF Gly 150	RMSF Asp 40
		RMSF Gly 124	RMSF Glu 107	RMSF Gly 124	RMSF Gly 150
		RMSF Asp 53	RMSF Val 169	RMSF Arg 176	RMSF Leu 120
		RMSF Glu 173	RMSF Asp 166	RMSF Asp 53	RMSF Tyr 34
		RMSF Ser 185	RMSF Val 73	RMSF Tyr 34	RMSF Thr 167
	Physicochemical model	RMSF Asp 158	RMSF His 82	RMSF Ala 165	RMSF Ala 98
		DEPR Leu 4	DEPR Val 139	DEPR Leu 4	DEPR His 51
		DEPR His 51	DEPR Thr 153	DEPR Phe 50	DEPR Thr 160
		DEPR Gln 141	DEPR Val 128	DEPR Thr 160	DEPR Gly 162
		DEPR Phe 106	DEPR Ser 179	DEPR Ala 79	DEPR Thr 153
		DEPR Leu 182	DEPR His 71	DEPR Thr 142	DEPR Thr 142
	Geometrical model	DEPR Gly 162	DEPR Val 147	DEPR Ser 104	DEPR Tyr 41
		B_I (AS) Leu 4	C_C138 Ser 74	B_I (AS) Leu 4	B_I (AS) Thr 142
		B_I (AS) Thr 142	B_I (AS) Leu 4	B_I (AS) Thr 142	P_C (MCS) Gly 97
		C_C138 Ala 161	B_P (MCS) Thr 160	C_C138 Val 157	B_I (AS) Leu 4
		I_P (MCS) Ser 65	C_K96 Ala 25	C_C138 Ala 102	C_C138 Ala 161
		C_D49 Val 131	C_D49 Asp 145	P_T (MCS) Leu 172	I_P (MCS) Ser 65
		I_C (MCS) Thr 160	B_I (AS) Val 139	I_P (MCS) Phe 50	C_C138 Val 157

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_α 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_{α} 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,

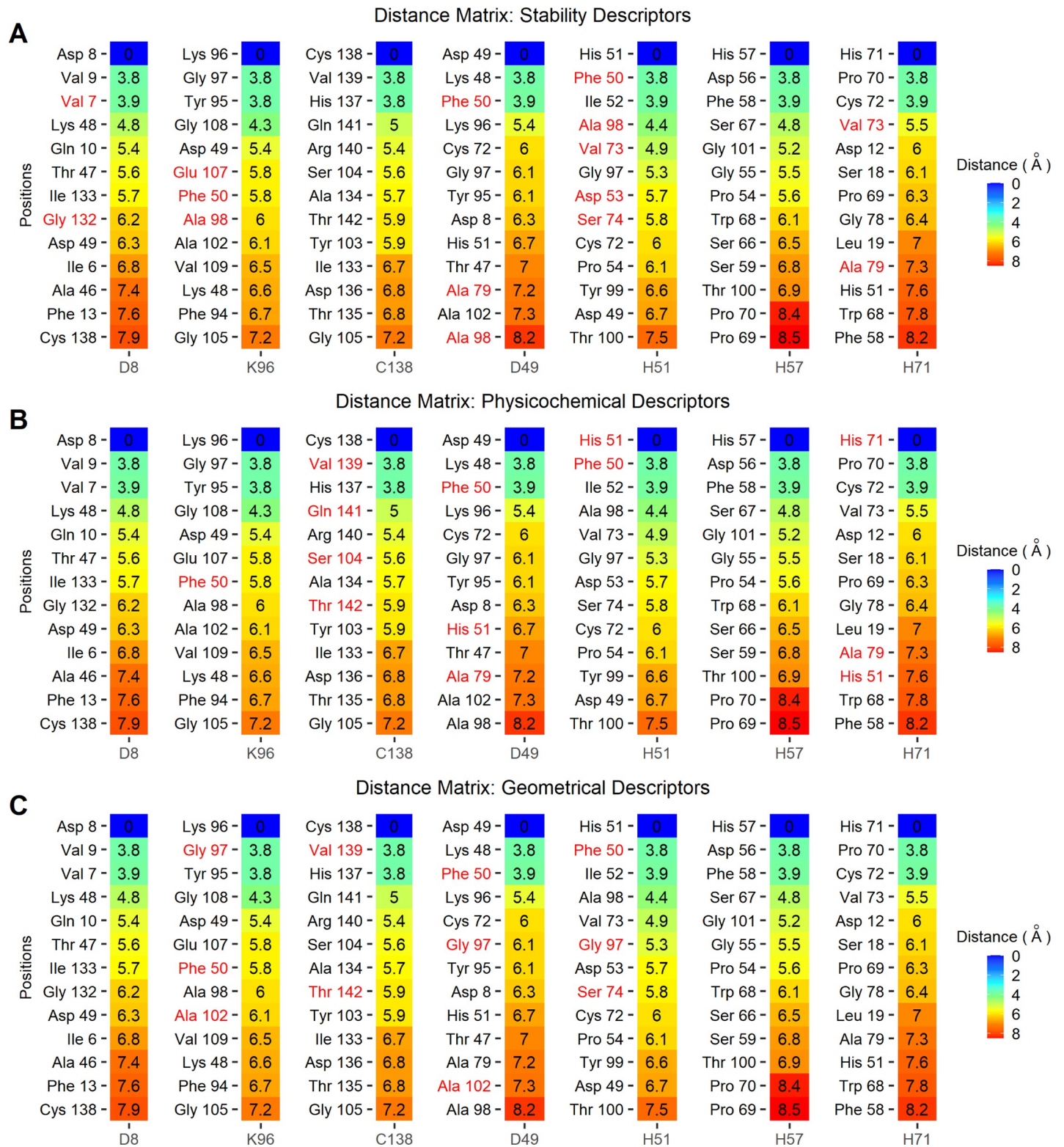


Fig 3. Distance matrix of WT-PZAse residues position against AS and MCS. The distance between pairs of C_{α} is sorted ascendingly. The matrix shows the 13 positions closest to the residues of the AS and MCS. The residues whose position is included in the selected descriptors are depicted in red for (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 PZAases 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 dichotomous 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.

2. Floyd K, Ph D, Jaramillo E, Ph D, Lönnroth K, Ph D, et al. MDR Tuberculosis—Critical Steps for Prevention and Control. *N Engl J Med*. 2010; 363: 1050–1058. <https://doi.org/10.1056/NEJMra0908076> PMID: 20825317
3. Zhang Y, Wade MM, Scorpio A, Zhang H, Sun Z. Mode of action of pyrazinamide: disruption of *Mycobacterium tuberculosis* membrane transport and energetics by pyrazinoic acid. 2017; 52: 790–795. <https://doi.org/10.1093/jac/dkg446> PMID: 14563891
4. Yang TW, Park HO, Jang HN, Yang JH, Kim SH, Moon SH, et al. Side effects associated with the treatment of multidrug-resistant tuberculosis at a tuberculosis referral hospital in South Korea. *Med (United States)*. 2017; 96: e7482. <https://doi.org/10.1097/MD.0000000000007482> PMID: 28700490
5. Vásquez-Campos L, Asencios-Solis L, Leo-Hurtado E, Quispe-Torres N, Salazar-Lindo E, Bayona J, et al. Drug resistance trends among previously treated tuberculosis patients in a national registry in Peru, 1994–2001. *Int J Tuberc Lung Dis*. 2004; 8: 465–472. PMID: 15141740
6. Whitfield MG, Soeters HM, Warren RM, York T, Sampson SL, Streicher EM, et al. A global perspective on pyrazinamide resistance: Systematic review and meta-analysis. *PLoS One*. 2015; 10: e0133869. <https://doi.org/10.1371/journal.pone.0133869> PMID: 26218737
7. 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*. 2009; 89: 109–113. <https://doi.org/10.1016/j.tube.2009.01.004> PMID: 19249243
8. Zhang Y, Mitchison D, Shi W, Zhang W. Mechanisms of Pyrazinamide Action and Resistance. *Microbiol Spectr*. 2014; 2: MGM2–0023. <https://doi.org/10.1128/microbiolspec.MGM2-0023-2013> PMID: 25530919
9. Cheng SJ, Thibert L, Sanchez T, Heifets L, Zhang Y. pncA mutations as a major mechanism of pyrazinamide resistance in *Mycobacterium tuberculosis*: Spread of a mono-resistant strain in Quebec, Canada. *Antimicrob Agents Chemother*. 2000; 44: 528–532. <https://doi.org/10.1128/aac.44.3.528-532.2000> PMID: 10681313
10. 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*. 2017; 23: 491–495. <https://doi.org/10.3201/eid2303.161034> PMID: 28221108
11. Mestdagh M, Fonteyne PA, Realini L, Rossau R, Jannes G, Mijs W, et al. Relationship between pyrazinamide resistance, loss of pyrazinamidase activity, and mutations in the pncA locus in multidrug-resistant clinical isolates of *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother*. 1999; 43: 2317–2319. <https://doi.org/10.1128/AAC.43.9.2317> PMID: 10471589
12. Sreevatsan S, Pan X, Zhang Y, Kreiswirth BN, Musser JM. Mutations associated with pyrazinamide resistance in pncA of *Mycobacterium tuberculosis* complex organisms. *Antimicrob Agents Chemother*. 1997; 41: 636–640. <https://doi.org/10.1128/AAC.41.3.636> PMID: 9056006
13. 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*. 2011; 6: e15785. <https://doi.org/10.1371/journal.pone.0015785> PMID: 21283666
14. Wabale VR, Joshi AA, Muthaiah M, Chowdhary AS. Wayne's Assay: A Screening Method for Indirect Detection of Pyrazinamide Resistance in *Mycobacterium tuberculosis* Clinical Isolates. *Bacteriol Microol*. 2017; 5: 1–9. <https://doi.org/10.15406/jbmoa.2017.05.00123>
15. Scorpio A, Lindholm-Levy P, Heifets Leonid, Gilman R, Siddiqi S, Cynamon M, et al. Characterization of pncA mutations in pyrazinamide-resistant *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother*. 1997; 41: 540–543. <http://dx.doi.org/10.3346/jkms.2001.16.5.537>
16. Khan MT, Malik SI, Ali S, Masood N, Nadeem T, Khan AS, et al. Pyrazinamide resistance and mutations in pncA among isolates of *Mycobacterium tuberculosis* from Khyber Pakhtunkhwa, Pakistan. *BMC Infect Dis*. 2019; 19: 116. <https://doi.org/10.1186/s12879-019-3764-2> PMID: 30728001
17. Caviedes L, Moore D. Introducing mods: A low-cost, low-tech tool for high-performance detection of tuberculosis and multidrug resistant tuberculosis. *Indian Journal of Medical Microbiology*. 2007. p. 87. <https://doi.org/10.4103/0255-0857.32711> PMID: 17582175
18. Alcántara R, Fuentes P, Marin L, Kirwan DE, Gilman RH, Zimic M, et al. Direct determination of pyrazinamide (PZA) susceptibility by sputum microscopic observation drug susceptibility (MODS) culture at neutral pH: The MODS-PZA assay. *J Clin Microbiol*. 2020; 58: e01165–19. <https://doi.org/10.1128/JCM.01165-19> PMID: 32132191
19. Aono A, Chikamatsu K, Yamada H, Igarashi Y, Murase Y, Takaki A, et al. A simplified pyrazinamidase test for pyrazinamide drug susceptibility in *Mycobacterium tuberculosis*. *J Microbiol Methods*. 2018; 154: 52–54. <https://doi.org/10.1016/j.mimet.2018.09.018> PMID: 30316980
20. Alcántara R, Fuentes P, Antiparra R, Santos M, Gilman RH, Kirwan DE, et al. MODS-Wayne, a colorimetric adaptation of the Microscopic-Observation Drug Susceptibility (MODS) assay for detection of

- mycobacterium tuberculosis pyrazinamide resistance from sputum samples. *J Clin Microbiol*. 2019. <https://doi.org/10.1128/JCM.01162-18> PMID: 30429257
21. Martin A, Takiff H, Vandamme P, Swings J, Palomino JC, Portaels F. A new rapid and simple colorimetric method to detect pyrazinamide resistance in *Mycobacterium tuberculosis* using nicotinamide. *J Antimicrob Chemother*. 2006; 58: 327–331. <https://doi.org/10.1093/jac/dkl231> PMID: 16751203
 22. Maningi NE, Daum LT, Rodriguez JD, Mphahlele M, Peters RPH, Fischer GW, et al. Improved detection by next-generation sequencing of pyrazinamide resistance in *Mycobacterium tuberculosis* isolates. *J Clin Microbiol*. 2015; 53: 3779–3783. <https://doi.org/10.1128/JCM.01179-15> PMID: 26378284
 23. Phelan J, O'Sullivan DM, Machado D, Ramos J, Whale AS, O'Grady J, et al. The variability and reproducibility of whole genome sequencing technology for detecting resistance to anti-tuberculous drugs. *Genome Med*. 2016; 8: 132. <https://doi.org/10.1186/s13073-016-0385-x> PMID: 28003022
 24. Madrazo-Moya CF, Cancino-Muñoz I, Cuevas-Córdoba B, González-Covarrubias V, Barbosa-Amez-cua M, Soberón X, et al. Whole genomic sequencing as a tool for diagnosis of drug and multidrug-resistance tuberculosis in an endemic region in Mexico. *PLoS One*. 2019; 14: e0213046. <https://doi.org/10.1371/journal.pone.0213046> PMID: 31166945
 25. Yadon AN, Maharaj K, Adamson JH, Lai YP, Sacchettini JC, Ioerger TR, et al. A comprehensive characterization of PncA polymorphisms that confer resistance to pyrazinamide. *Nat Commun*. 2017; 8: 588. <https://doi.org/10.1038/s41467-017-00721-2> PMID: 28928454
 26. Wu X, Lu W, Shao Y, Song H, Li G, Li Y, et al. pncA gene mutations in reporting pyrazinamide resistance among the MDR-TB suspects. *Infect Genet Evol*. 2019; 72: 147–150. <https://doi.org/10.1016/j.meegid.2018.11.012> PMID: 30447296
 27. Sengstake S, Bergval IL, Schuitema AR, de Beer JL, Phelan J, de Zwaan R, et al. Pyrazinamide resistance-conferring mutations in pncA and the transmission of multidrug resistant TB in Georgia. *BMC Infect Dis*. 2017; 17. <https://doi.org/10.1186/s12879-017-2594-3> PMID: 28697808
 28. Masso M, Vaisman II. Structure-based prediction of protein activity changes: Assessing the impact of single residue replacements. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. 2011. pp. 3221–3224. <https://doi.org/10.1109/IEMBS.2011.6090876> PMID: 22255025
 29. Cirillo DM, Miotto P, Niemann S. Use of genetic mutations for prediction of resistance to pyrazinamide in *M. tuberculosis* strains. *Int J Mycobacteriology*. 2015; 4: 42–43. <https://doi.org/10.1080/10893950390219083>
 30. Berezovsky IN, Guarnera E, Zheng Z, Eisenhaber B, Eisenhaber F. Protein function machinery: from basic structural units to modulation of activity. *Curr Opin Struct Biol*. 2017; 42: 67–74. <https://doi.org/10.1016/j.sbi.2016.10.021> PMID: 27865209
 31. Masso M, Vaisman II. A structure-based computational mutagenesis elucidates the spectrum of stability-activity relationships in proteins. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. 2011. pp. 3225–8. <https://doi.org/10.1109/IEMBS.2011.6090877> PMID: 22255026
 32. Chen C-Y, Georgiev I, Anderson AC, Donald BR. Computational structure-based redesign of enzyme activity. *Proc Natl Acad Sci*. 2009; 106: 3764–3769. <https://doi.org/10.1073/pnas.0900266106> PMID: 19228942
 33. Karmakar M, Rodrigues CHM, Horan K, Denholm JT, Ascher DB. Structure guided prediction of Pyrazinamide resistance mutations in pncA. *Sci Rep*. 2020;10. <https://doi.org/10.1038/s41598-020-58635-x> PMID: 32024884
 34. Gherardini PF, Helmer-Citterich M. Structure-based function prediction: Approaches and applications. *Briefings in Functional Genomics and Proteomics*. 2008. <https://doi.org/10.1093/bfgp/eln030> PMID: 18599513
 35. Mehmood A, Jones SI, Tao P, Janesko BG. An Orbital-Overlap Complement to Ligand and Binding Site Electrostatic Potential Maps. *J Chem Inf Model*. 2018; 58: 1836–1846. <https://doi.org/10.1021/acs.jcim.8b00370> PMID: 30160959
 36. Sivanandam M, Saravanan K, Kumaradhas P. Insights into intermolecular interactions, electrostatic properties and the stability of C646 in the binding pocket of p300 histone acetyltransferase enzyme: a combined molecular dynamics and charge density study. *J Biomol Struct Dyn*. 2017; 36: 3246–3264. <https://doi.org/10.1080/07391102.2017.1384761> PMID: 28948877
 37. Ruiz-Pernía JJ, Martí S, Moliner V, Tuñón I. A novel strategy to study electrostatic effects in chemical reactions: Differences between the role of solvent and the active site of chalcone isomerase in a michael addition. *J Chem Theory Comput*. 2012; 8: 1532–1535. <https://doi.org/10.1021/ct300064f> PMID: 26593647

38. Quiliano M, Gutierrez AH, Gilman RH, López C, Evangelista W, Sotelo J, et al. Structure-Activity relationship in mutated pyrazinamidases from *Mycobacterium tuberculosis*. *Bioinformatics*. 2011; 6: 335–339. <https://doi.org/10.6026/97320630006335> PMID: 21814390
39. Sacquin-Mora S. Bridging Enzymatic Structure Function via Mechanics: A Coarse-Grain Approach. *Methods Enzymol*. 2016; 578: 227–248. <https://doi.org/10.1016/bs.mie.2016.05.022> PMID: 27497169
40. Hvidsten TR, Lægread A, Kryshafovych A, Anderson G, Fidelis K, Komorowski J. A comprehensive analysis of the structure-function relationship in proteins based on local structure similarity. *PLoS One*. 2009; 4: e6266. <https://doi.org/10.1371/journal.pone.0006266> PMID: 19603073
41. Ng PC, Henikoff S. Predicting the Effects of Amino Acid Substitutions on Protein Function. *Annu Rev Genomics Hum Genet*. 2006; 7: 61–68. <https://doi.org/10.1146/annurev.genom.7.080505.115630> PMID: 16824020
42. Bromberg Y, Rost B. Correlating protein function and stability through the analysis of single amino acid substitutions. *BMC Bioinformatics*. 2009; 10: S8. [https://doi.org/10.1016/0022-0248\(78\)90328-7](https://doi.org/10.1016/0022-0248(78)90328-7)
43. Pazhang M, Mardi N, Mehrnejad F, Chaparzadeh N. The combinatorial effects of osmolytes and alcohols on the stability of pyrazinamidase: Methanol affects the enzyme stability through hydrophobic interactions and hydrogen bonds. *Int J Biol Macromol*. 2018; 108: 1339–1347. <https://doi.org/10.1016/j.ijbiomac.2017.11.039> PMID: 29129628
44. Karshikoff A, Nilsson L, Ladenstein R. Rigidity versus flexibility: The dilemma of understanding protein thermal stability. *FEBS J*. 2015; 282: 3899–3917. <https://doi.org/10.1111/febs.13343> PMID: 26074325
45. Esmaeeli R, Mehrnejad F, Mir-Derikvand M, Gopalpoor N. Computational insights into pH-dependence of structure and dynamics of pyrazinamidase: A comparison of wild type and mutants. *J Cell Biochem*. 2018. <https://doi.org/10.1002/jcb.27543> PMID: 30304542
46. Khan MT, Junaid M, Mao X, Wang Y, Hussain A, Malik SI, et al. Pyrazinamide resistance and mutations L19R, R140H, and E144K in Pyrazinamidase of *Mycobacterium tuberculosis*. *J Cell Biochem*. 2018. <https://doi.org/10.1002/jcb.27989> PMID: 30485476
47. Khan MT, Malik SI. Structural dynamics behind variants in pyrazinamidase and pyrazinamide resistance. *J Biomol Struct Dyn*. 2019; 1–15. <https://doi.org/10.1080/07391102.2019.1650113> PMID: 31357912
48. Junaid M, Khan MT, Malik SI, Wei D-Q. Insights into the Mechanisms of the Pyrazinamide Resistance of Three Pyrazinamidase Mutants N11K, P69T, and D126N. *J Chem Inf Model*. 2018; acs.jcim.8b00525. <https://doi.org/10.1021/acs.jcim.8b00525> PMID: 30481017
49. Aggarwal M, Singh A, Grover S, Pandey B, Kumari A, Grover A. Role of *pncA* gene mutations W68R and W68G in pyrazinamide resistance. *J Cell Biochem*. 2018; 119: 2567–78. <https://doi.org/10.1002/jcb.26420> PMID: 28980723
50. Tidor B, Karplus M. The contribution of cross-links to protein stability: A normal mode analysis of the configurational entropy of the native state. *Proteins Struct Funct Bioinforma*. 1993; 15: 71–79. <https://doi.org/10.1002/prot.340150109> PMID: 7680808
51. Bauer JA, Pavlovic J, Bauerová-Hlinková V. Normal mode analysis as a routine part of a structural investigation. *Molecules*. 2019; 23: 3293. <https://doi.org/10.3390/molecules24183293> PMID: 31510014
52. Skjaerven L, Hollup SM, Reuter N. Normal mode analysis for proteins. *J Mol Struct THEOCHEM*. 2009; 898: 42–48. <https://doi.org/10.1016/j.theochem.2008.09.024>
53. Skjaerven L, Martinez A, Reuter N. Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit. *Proteins Struct Funct Bioinforma*. 2011; 79: 232–243. <https://doi.org/10.1002/prot.22875> PMID: 21058295
54. López-Blanco JR, Miyashita O, Tama F, Chacón P. Normal Mode Analysis Techniques in Structural Biology. *eLS*. 2014. <https://doi.org/10.1002/9780470015902.a0020204.pub2>
55. Tama F, Gadea FX, Marques O, Sanejouand YH. Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins Struct Funct Genet*. 2000; 41: 1–7. [https://doi.org/10.1002/1097-0134\(20001001\)41:1<1::aid-prot10>3.0.co;2-p](https://doi.org/10.1002/1097-0134(20001001)41:1<1::aid-prot10>3.0.co;2-p) PMID: 10944387
56. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018. <https://doi.org/10.1093/nar/gky427> PMID: 29788355
57. López-Blanco JR, Aliaga JI, Quintana-Ortí ES, Chacón P. IMODS: Internal coordinates normal mode analysis server. *Nucleic Acids Res*. 2014; 42: W271–W276. <https://doi.org/10.1093/nar/gku339> PMID: 24771341
58. 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*. 2015; 1–2: 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>

59. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc.* 1996; 118: 11225–11236. <https://doi.org/10.1021/ja9621760>
60. Berendsen HJC, Grigera JR, Straatsma TP. The missing term in effective pair potentials. *J Phys Chem.* 1987; 91: 6269–6271. <https://doi.org/10.1021/j100308a038>
61. Berendsen HJC. Transport Properties Computed by Linear Response through Weak Coupling to a Bath. *Computer Simulation in Materials Science.* 1991. pp. 139–155. https://doi.org/10.1007/978-94-011-3546-7_7
62. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys.* 1981; 52. <https://doi.org/10.1063/1.328693>
63. Nosé S, Klein ML. Constant pressure molecular dynamics for molecular systems. *Mol Phys.* 1983; 50. <https://doi.org/10.1080/00268978300102851>
64. Hockney RW, Goel SP, Eastwood JW. Quiet high-resolution computer models of a plasma. *J Comput Phys.* 1974; 14: 148–158. [https://doi.org/10.1016/0021-9991\(74\)90010-2](https://doi.org/10.1016/0021-9991(74)90010-2)
65. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A Linear Constraint Solver for molecular simulations. *J Comput Chem.* 1997; 18: 1463–1472. [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H)
66. Ewald PP. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann Phys.* 1921; 369: 253–287. <https://doi.org/10.1002/andp.19213690304>
67. Hunter JD. Matplotlib: A 2D graphics environment. *Comput Sci Eng.* 2007; 9: 90–95. <https://doi.org/10.1109/MCSE.2007.55>
68. Valdebenito-Maturana B, Reyes-Suarez JA, Henriquez J, Holmes DS, Quatrini R, Pohl E, et al. Mutantelec: An In Silico mutation simulation platform for comparative electrostatic potential profiling of proteins. *J Comput Chem.* 2017; 38: 467–474. <https://doi.org/10.1002/jcc.24712> PMID: 28114729
69. Malmberg CG, Maryott AA. Dielectric constant of water from 0 to 100 C. *J Res Natl Bur Stand (1934).* 1956; 56: 1–8. <https://doi.org/10.6028/jres.056.001>
70. Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, et al. The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry.* 2011; 50: 4402–4410. <https://doi.org/10.1021/bi2002289> PMID: 21506553
71. Kuhn M. Building Predictive Models in R Using the caret Package. *J Stat Softw.* 2008; 28: 1–26. <https://doi.org/10.18637/jss.v028.i07> PMID: 27774042
72. Bennett G. The merck index: An encyclopedia of chemicals, drugs and biologicals. *J Hazard Mater.* 1992. [https://doi.org/10.1016/0304-3894\(92\)87022-8](https://doi.org/10.1016/0304-3894(92)87022-8)
73. Doustdar F, Pazhang M, Mehrnejad F, Safarzadeh M, Rabiei D, Chaparzadeh N, et al. Biochemical Characterization and Computational Identification of *Mycobacterium tuberculosis* Pyrazinamidase in Some Pyrazinamide-Resistant Isolates of Iran. *Protein J.* 2015; 34: 181–192. <https://doi.org/10.1007/s10930-015-9610-5> PMID: 25972249
74. Vats C, Dhanjal JK, Goyal S, Gupta A, Bharadvaja N, Grover A. Mechanistic analysis elucidating the relationship between Lys96 mutation in *Mycobacterium tuberculosis* pyrazinamidase enzyme and pyrazinamide susceptibility. *BMC Genomics.* 2015; 16. <https://doi.org/10.1186/1471-2164-16-S2-S14> PMID: 25708048
75. Whitfield MG, Warren RM, Streicher EM, Sampson SL, Sirgel FA, Van Helden PD, et al. *Mycobacterium tuberculosis* pncA polymorphisms that do not confer pyrazinamide resistance at a breakpoint concentration of 100 micrograms per milliliter in MGIT. *J Clin Microbiol.* 2015; 53: 3633–3635. <https://doi.org/10.1128/JCM.01001-15> PMID: 26292310
76. Yoon J-H, Nam J-S, Kim K-J, Ro Y-T. Characterization of pncA mutations in pyrazinamide-resistant *Mycobacterium tuberculosis* isolates from Korea and analysis of the correlation between the mutations and pyrazinamidase activity. *World J Microbiol Biotechnol.* 2014; 30: 2821–2828. <https://doi.org/10.1007/s11274-014-1706-0> PMID: 25034468
77. Fuentes P, Pesaresi A, Gilman RH, Gutiérrez AH, Olivera P, Evangelista W, et al. Role of Metal Ions on the Activity of *Mycobacterium tuberculosis* Pyrazinamidase. *Am J Trop Med Hyg.* 2012; 87: 153–161. <https://doi.org/10.4269/ajtmh.2012.10-0565> PMID: 22764307
78. Zhang H, Deng JY, Bi LJ, Zhou YF, Zhang ZP, Zhang CG, et al. Characterization of *Mycobacterium tuberculosis* nicotinamidase/ pyrazinamidase. *FEBS J.* 2008; 275: 753–62. <https://doi.org/10.1111/j.1742-4658.2007.06241.x> PMID: 18201201
79. Khadem-Maaref M, Mehrnejad F, Phirouznia A. Effects of metal-ion replacement on pyrazinamidase activity: A quantum mechanical study. *J Mol Graph Model.* 2017; 73: 24–29. <https://doi.org/10.1016/j.jmgl.2017.01.021> PMID: 28214629

80. Rasool N, Hussain W, Khan YD. Revelation of enzyme activity of mutant pyrazinamidases from *Mycobacterium tuberculosis* upon binding with various metals using quantum mechanical approach. *Comput Biol Chem*. 2019; 83: 107108. <https://doi.org/10.1016/j.compbiolchem.2019.107108> PMID: 31442707
81. Carter JJ, Walker TM, Walker AS, Whitfield MG, Morlock GP, Peto TEA, et al. Prediction of Pyrazinamide Resistance in *Mycobacterium Tuberculosis* Using Structure-Based Machine Learning Approaches. *SSRN Electron J*. 2019. <https://doi.org/10.2139/ssrn.3391941>
82. Tam KKG, Leung KSS, Siu GKH, Chang KC, Wong SSSY, Ho PL, et al. Direct Detection of Pyrazinamide Resistance in *Mycobacterium tuberculosis* by Use of *pncA* PCR Sequencing. *J Clin Microbiol*. 2019. <https://doi.org/10.1128/JCM.00145-19> PMID: 31189582
83. Maslov DA, Zaichikova M V., Chernousova LN, Shur K V., Bekker OB, Smirnova TG, et al. Resistance to pyrazinamide in Russian *Mycobacterium tuberculosis* isolates: *PncA* sequencing versus Bactec MGIT 960. *Tuberculosis*. 2015; 95: 608–612. <https://doi.org/10.1016/j.tube.2015.05.013> PMID: 26071666
84. Sheen P, Lozano K, Gilman RH, Valencia HJ, Loli S, Fuentes P, et al. *PncA* gene expression and prediction factors on pyrazinamide resistance in *Mycobacterium tuberculosis*. *Tuberculosis*. 2013; 93: 515–522. <https://doi.org/10.1016/j.tube.2013.03.005> PMID: 23867321
85. Sun Q, Li X, Perez LM, Shi W, Zhang Y, Sacchetti JC. The molecular basis of pyrazinamide activity on *Mycobacterium tuberculosis* PanD. *Nat Commun*. 2020; 11. <https://doi.org/10.1038/s41467-019-14238-3> PMID: 31953389
86. Lamont EA, Baughn AD. Impact of the host environment on the antitubercular action of pyrazinamide. *EBioMedicine*. 2019. pp. 374–380. <https://doi.org/10.1016/j.ebiom.2019.10.014> PMID: 31669220
87. Werngren J, Alm E, Mansjö M. Non-*pncA* gene-mutated but pyrazinamide-resistant *Mycobacterium tuberculosis*: Why is that? *J Clin Microbiol*. 2017; 55: 1920–1927. <https://doi.org/10.1128/JCM.02532-16> PMID: 28404681
88. 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*. 2014; 58: 4928–4930. <https://doi.org/10.1128/AAC.02394-14> PMID: 24867972
89. Hunt M, Bradley P, Lapierre SG, Heys S, Thomsit M, Hall MB, et al. Antibiotic resistance prediction for *Mycobacterium tuberculosis* from genome sequence data with Mykrobe. *Wellcome Open Res*. 2019. <https://doi.org/10.12688/wellcomeopenres.15603.1> PMID: 32055708
90. Chaidir L, Ruesen C, Dutilh BE, Ganiem AR, Andryani A, Apriani L, et al. Use of whole-genome sequencing to predict *Mycobacterium tuberculosis* drug resistance in Indonesia. *J Glob Antimicrob Resist*. 2019; 16: 170–177. <https://doi.org/10.1016/j.jgar.2018.08.018> PMID: 30172045
91. Schleusener V, Köser CU, Beckert P, Niemann S, Feuerriegel S. *Mycobacterium tuberculosis* resistance prediction and lineage classification from genome sequencing: Comparison of automated analysis tools. *Sci Rep*. 2017; 7. <https://doi.org/10.1038/srep46327> PMID: 28425484
92. Iwamoto T, Murase Y, Yoshida S, Aono A, Kuroda M, Sekizuka T, et al. Overcoming the pitfalls of automatic interpretation of whole genome sequencing data by online tools for the prediction of pyrazinamide resistance in *Mycobacterium tuberculosis*. *PLoS One*. 2019; 14: e0212798. <https://doi.org/10.1371/journal.pone.0212798> PMID: 30817803