



Physiological modeling of the metaverse of the *Mycobacterium tuberculosis* β -CA inhibition mechanism

Simone Giovannuzzi^{a,1}, Sagar Singh Shyamal^{b,1}, Ratul Bhowmik^{c,1}, Rajarshi Ray^c,
Ajay Manaithiya^c, Fabrizio Carta^a, Seppo Parrkila^{c,d}, Ashok Aspatwar^{c,*}, Claudiu T. Supuran^a

^a Department of Neuroscience, Psychology, Drug Research, and Child's Health, Section of Pharmaceutical and Nutraceutical Sciences, University of Florence, Via Ugo Schiff 6, 50019, Sesto Fiorentino, Italy

^b Department of Pharmaceutical Engineering & Technology, Indian Institute of Technology (Banaras Hindu University), Varanasi, India

^c Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

^d Fimlab Ltd, Tampere University Hospital, Tampere, Finland

ARTICLE INFO

Keywords:

Mechanistic interpretation modeling
Rational synthesis
Carbonic anhydrase
Mycobacterium tuberculosis
Infection
QSAR
Systems biology
Mechanism of action
Immunity
Molecular docking

ABSTRACT

Tuberculosis (TB) is an infectious disease that primarily affects the lungs of humans and accounts for *Mycobacterium tuberculosis* (Mtb) bacteria as the etiologic agent. In this study, we introduce a computational framework designed to identify the important chemical features crucial for the effective inhibition of Mtb β -CAs. Through applying a mechanistic model, we elucidated the essential features pivotal for robust inhibition. Using this model, we engineered molecules that exhibit potent inhibitory activity and introduce relevant novel chemistry. The designed molecules were prioritized for synthesis based on their predicted pKi values via the QSAR (Quantitative Structure-Activity Relationship) model. All the rationally designed and synthesized compounds were evaluated *in vitro* against different carbonic anhydrase isoforms expressed from the pathogen Mtb; moreover, the off-target and widely human-expressed CA I and II were also evaluated. Among the reported derivatives, 2, 4, and 5 demonstrated the most valuable *in vitro* activity, resulting in promising candidates for the treatment of TB infection. All the synthesized molecules exhibited favorable pharmacokinetic and toxicological profiles based on *in silico* predictions. Docking analysis confirmed that the zinc-binding groups bind effectively into the catalytic triad of the Mtb β -Cas, supporting the *in vitro* outcomes with these binding interactions. Furthermore, molecules with good prediction accuracies according to previously established mechanistic and QSAR models were utilized to delve deeper into the realm of systems biology to understand their mechanism in combating tuberculous pathogenesis. The results pointed to the key involvement of the compounds in modulating immune responses via NF- κ B1, SRC kinase, and TNF- α to modulate granuloma formation and clearance via T cells. This dual action, in which the pathogen's enzyme is inhibited while modulating the human immune machinery, represents a paradigm shift toward more effective and comprehensive treatment approaches for combating tuberculosis.

1. Introduction

Global estimates from the World Health Organization (WHO) in 2021 show more than 10 million people acquired tuberculosis (TB) infections caused by *Mycobacterium tuberculosis* (Mtb) or related mycobacterial strains [1], and a consistent number of affected patients developed serious illness, leading to fatal outcomes [2]. Although TB is among the oldest known human-affecting diseases, the actual therapeutic regimen involves a combination of outdated antibiotics such as

isoniazid, rifampin, pyrazinamide, ethambutol, streptomycin, fluoroquinolones, amikacin, kanamycin, and capreomycin over for at least six months [3]. (Fig. 1).

Countries with populations particularly exposed to an elevated incidence of TB infections, such as infants and children, are very likely to use the bacillus Calmette-Guérin (BCG) vaccine for prevention purposes. New advancements in TB-applied vaccines, such as VPM1002, M72/AS01E, MTBVAC, Ad5Ag85A, and RUTI, are the most promising, although definitive therapeutic evidences have yet to be reported [4].

* Corresponding author.

E-mail addresses: ashok.aspatwar@tuni.fi (A. Aspatwar), claudiu.supuran@unifi.it (C.T. Supuran).

¹ These authors contributed equally to this work.

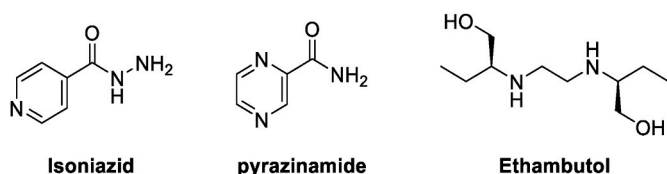


Fig. 1. Chemical structures of several TB drugs used.

Despite intense research efforts in TB therapeutics, the disease remains far from eradicated, mainly because of the interaction of the pathogen with the host tissues. *Mtb* allocates within the lungs and strongly activates the host immune system by recruiting macrophages and lymphocytes, which trigger a robust cell-mediated immune response. As a result, *Mtbs* remain physically seized within a structure formed by an amorphous collection of macrophages and other immune cells called granulomas. In immunocompetent individuals, granulomas cannot eliminate the pathogen but restrain the bacteria and halt the progression of the disease (i.e., TB). This process establishes a hospitable niche for *Mtbs*, allowing them to survive for decades by being hidden from the immune system and persisting in a non or slowly replicating state, which is highly beneficial for the development of multidrug-resistant (MDR) and extensively drug-resistant (XDR) *Mtb* strain [5,6].

In such a context, the validation of new druggable targets represents the ideal answer to the management of MDR events [7–9]. Convincing evidence sustains the carbonic anhydrase (CA, EC 4.2.1.1) enzymes expressed by the *Mtb* pathogen as druggable targets [10–12]. Specifically, *Mtb* encodes three β -CAs, Rv1284 (*Mtb* CA1), Rv3588c (*Mtb* CA2) and Rv3273 (*Mtb* CA3), all of which play crucial physiological roles in

the bacterial life cycle [12–14] invasion and survival within the human host environment [11].

Human carbonic anhydrases (hCAs) also play various roles in physiological processes such as CO_2 and bicarbonate transport, respiration, pH balancing, CO_2 homeostasis, electrolyte secretion, and biosynthetic reactions [15]. Distinct, evolutionarily unrelated gene families of CAs are present in various organisms, with the α -class present in humans as 15 different isoforms (hCA I–XIV) [16], hCA I is found in red blood cells and many tissues, and while its exact physiological function is unknown, it is associated with conditions such as retinal and cerebral edema, where its inhibition may be beneficial. hCA II, the physiologically dominant isoform, is associated with several disease conditions, such as epilepsy, edema, glaucoma, and altitude sickness. Recently, hCAs have also emerged as potential targets for designing anti-infective drugs with novel mechanisms of action [17–20].

In vitro inhibition experiments have clearly shown that diverse CA inhibitors (CAIs) of primary sulfonamides as well as dithiocarbamate (DTC) types efficiently suppress *Mtb* β -CA activity (Fig. 2) [12–14, 21–28].

However, it is essential to design inhibitors that selectively target *Mtb* β -CAs over hCAs to avoid off-target effects that could harm the host. Thus, our research focuses on developing selective inhibitors for *Mtb* β -CAs while minimizing interactions with hCAs I and II.

In light of these results, we envisioned exploring new chemical entities as potential *Mtb* β -CAIs bearing both primary sulfonamide and DTC moieties using mechanistic interpretation models to guide the chemical ideation and incorporation of chemical motifs crucial for *Mtb* β -CAs inhibition. Through quantitative structural activity relationship (QSAR) models, we prioritized molecules to be synthesized. Synthesized

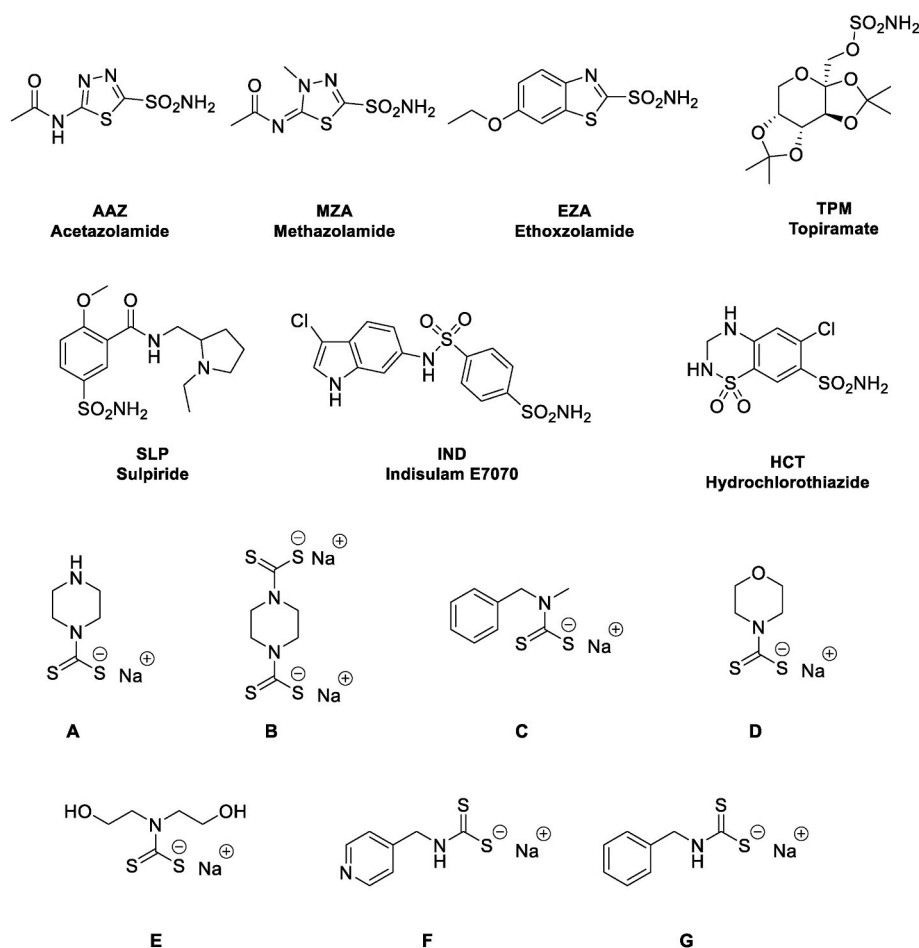


Fig. 2. Structures of primary sulfonamides (AAZ–HCT) and thiocarbamate moieties (A–G).

small molecules were profiled *in vitro* for their potency and selectivity against Mtb CA 1–3 over their off-target human (h)-expressed CAs I and II. After their biological evaluation, ADMET analysis was performed to ensure the compounds have favorable pharmacokinetics and a toxicological profile. To correlate the *in vitro* findings, docking studies were conducted to understand binding modes and molecular interactions, which are significant for their robust inhibition. Also, we deciphered alternative routes for the synthesized molecules to combat tuberculosis infection via systems biology approaches using different graph-based, statistical analysis, and hierarchical clustering approaches. Fig. 3 provides a bird's-eye view of the entire workflow employed in this study (Table 1 describes the summary of employed softwares and techniques used in this study).

2. Results and discussion

2.1. Mechanistic interpretation (Multinomial Classification) models

A mechanistic interpretation based on the available datasets from the ChEMBL database concerning Mtb β -CAs was conducted to synthesize molecules based on recurring patterns rationally. A multinomial classification model was constructed for this task, where each molecule was labeled based on the target class with the highest pKi value. For example, if the pKi value for Mtb β -CA3 was the highest among the three targets for a particular molecule, it was labeled Mtb β -CA3 (generated dataset matrix is given in supplementary 1, Table S1). Four machine learning (ML) models, namely, the naïve Bayes (NB), random forest (RF), gradient boosting (GB), and generalized linear (GL) models, were tested, among which the RF method outperformed the other models based on the low difference in prediction accuracy between the training (97.9 %) and test sets (92.2 %) and other evaluations matrices mentioned in Table 2 (Detailed Workflow for the mechanistic model development is depicted in Fig. 4a and b). The RF model was validated through a tenfold cross-validation loop (Table 3), where it demonstrated comparable mean accuracy to the GB model, but the RF model had a lower standard deviation such that the predictions between different distributions of the training set were low, thus ensuring more robust and consistent performance across varying datasets. The receiver operating

characteristic (ROC) plots, generated in a one-vs-all manner (Fig. 7a), demonstrated robust prediction accuracies for the Mtb β -CA1, β -CA2, and β -CA3 class predictions. The area under the curve (AUC) values were notably high, with Mtb β -CA1 achieving an AUC of 0.96, Mtb β -CA2 achieving a perfect AUC of 1.0, and Mtb β -CA3 achieving an AUC of 0.96. These results indicate the strong predictive performance of the model for these classes, with particularly excellent performance for Mtb β -CA2. Hence, the RF model was selected based on its excellent performance matrices and high prediction stability based on the cross-validation study.

The top thirty important variables from the RF model (supplementary 1, Table S2) were used to make mechanistic interpretations to understand the specificity of molecules toward the inhibition of Mtb β -CA1-3 based upon chemical and biological endpoints as descriptors. Correlation analysis of the top 30 important features and Mtb β -CA1-3 pKi values revealed that the chemical spaces of β -CA1 and β -CA3 are significantly similar, as indicated by high Pearson correlation coefficients (Fig. 5a). This similarity suggests that the active sites of β -CA1 and β -CA3 may have similarities, implying that designing inhibitors for one isoform could simultaneously target the other isoform.

One might expect that biological endpoints play a crucial role in determining a molecule's specificity for a given target. However, to our surprise, the second most important descriptor was the autocorrelation descriptor ATSC8e. This descriptor, which is an average-centered Broto-Moreau autocorrelation at lag 8 weighted by Sanderson electronegativity, highlights the significant role of electronegativity in Mtb β -CA inhibition. Many autocorrelation descriptors were also found in the top thirty descriptors, which included descriptors related to Charges (GATS4c, MATS3c, MATS1c), Van Der Waals volume (ATSC5v, ATSC1v, AATSC2v, ATSC2v), mass (GATS2m, MATS1m) and ionization potential and polarizabilities (MATS4i and AATSC8p). Structural count-based descriptors were also analyzed to understand the structural patterns recognized by the mechanistic interpretation model. The model suggested that the number of **double bonds**, **rotatable bonds**, **single bonds**, and **hydrogen bond donors** were positively correlated with Mtb β -CA1 and Mtb β -CA3 inhibition and that Mtb β -CA2 inhibition was only positively correlated with the number of **five-membered hetero rings** (n5 Heteroring). Interestingly, Mtb β -CA2 was negatively

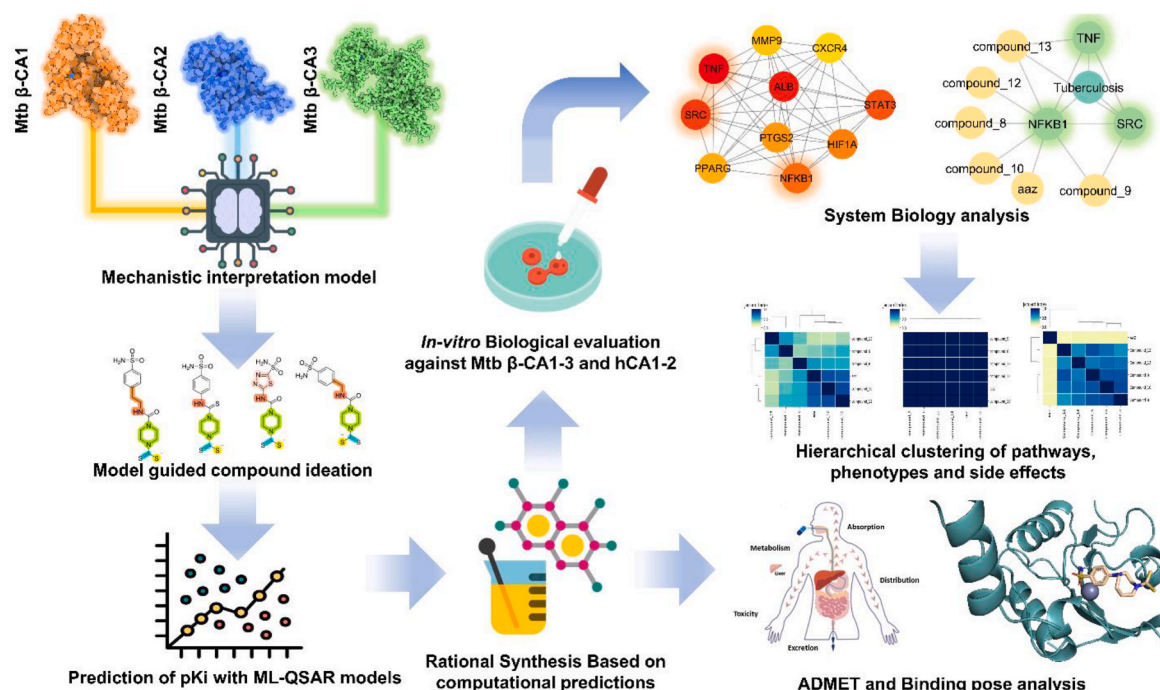


Fig. 3. Workflow of the study.

Table 1

Tabular representation of the computational tools and algorithms used in this study.

Tools/Methods	Background of Methodology	Motive of Use
PaDEL	Generation of chemical descriptors such as Topological, fingerprints, auto-correlation descriptors, etc.	To generate molecular descriptors for machine learning models.
H2O ML Nodes (KNIME)	Machine learning nodes	Creating and validating Mechanistic interpretation and ML-QSAR models
SwissADME	Webserver tool to predict pharmacokinetic and toxicity	ADME properties of synthesized molecules
Osiris Property explorer	Utilizes a weight-of-evidence (WoE) framework using the independent Bayesian approach	Toxicity prediction of the synthesized molecules
Autodock vina Zn	Molecular docking software for Zn containing enzymes	Docking of synthesized molecules on their respective metalloenzymes
Swiss Target prediction, SEA server, and PPB2	Webserver for predicting potential target genes for a chemical compound using varying methodologies	Collection of a diverse pool of potential targets of the synthesized molecules.
Comparative Toxicogenomic Database (CTD)	Cross-reference of data across biological databases and literature	To obtain biological data for respective diseases or biological conditions
VENNY 2.0	Set theory mathematics	To obtain similar genes among the considered biological conditions
David	Cross-reference of data across biological databases and literature	Perform pathway and GO term enrichment analysis
Correlation analysis for cluster determination	Jaccard Index and similarity matrix	To obtain a concentration of specific biological properties across chemical pairs
Correlation plots for hierarchical analysis	Neighbour-joining phylogenetic analysis	Similarity analysis of chemical pair w.r.t to their biological properties
Cytoscape	Graph-based modeling and node-edge interactions	Creation of complex biological networks
GTE portal	Cross-reference of data across biological databases and literature	Determining single-cell and organ-specific expression among the genes
NetInfer	Adverse event prediction via network inference method	Determination of potential side effects related to synthesized molecules

correlated with **Crippen's logP** descriptors. Structural fingerprints play an important role in model interpretation because they represent smaller distinct portions of a larger chemical structure, among which are two such fingerprints, **PubchemFP696** and **PubchemFP643**, which are related to the following smart patterns: **C-C-C-C-C-C-C** and **[#1]-C-C-N-[#1]**. In summary, this mechanistic interpretation study revealed many molecular descriptors and structural patterns. Fig. 5c and Table 4 depict these features that were incorporated into the planned synthetic molecules, such that the molecules were synthesized based on relevant novel chemistry and the identified structural motifs crucial for Mtb

Table 2

Evaluation of the Mechanistic interpretation model (classification).

ML_Set	overall Recall	overall Precision	overall sensitivity	overall Specificity	overall F-measure	Accuracy	Cohen's kappa
NB_Training	0.77	0.65	0.77	0.85	0.67	0.68	0.48
NB_Test	0.67	0.58	0.67	0.81	0.6	0.63	0.38
RF_Training	0.97	0.99	0.97	0.98	0.98	0.98	0.96
RF_Test	0.9	0.96	0.9	0.93	0.92	0.92	0.84
GB_Training	0.96	0.99	0.96	0.99	0.98	0.98	0.97
GB_Test	0.85	0.89	0.85	0.89	0.86	0.86	0.73
GL_Training	0.87	0.77	0.87	0.93	0.81	0.84	0.72
GL_Test	0.71	0.62	0.71	0.84	0.64	0.69	0.47

β -CAs inhibition.

2.2. ML-QSAR models for Mtb β -CAs

Before synthesis and *in vitro* evaluation, the strategically designed molecules, guided by our mechanistic interpretation models, were subjected to ML-QSAR models to predict their probable inhibition constant (pKi). For this task, four ML models were tested using the pre-processed set of descriptors.

For each class of target GB regression learner outperformed the other models (Evaluation matrices in Table 5). Models trained with the GB regression learner demonstrated very low differences in the predicted training and test class datasets, in which the robustness of the model was validated through tenfold cross-validation loops on the training sets (Table 6). Plots for the predicted pKi and experimental pKi values are shown in Fig. 7b, c, and 7d for all the GB regression models.

The feature importance of the variables and their corresponding correlation values are shown in Fig. 6, which includes many types of descriptors, such as electro-topological state atom type descriptors, autocorrelation descriptors (related to the van der Waals volume and charge), and information content descriptors (List of top 10 important from each model is given supplementary 1, Table S3).

Electrotopological state atom type descriptors, which combine the electronic and topological characteristics of atoms in a molecule, such as nHbint8 and nHbint6, were the top-ranking descriptors in the regression models of GB_Mt β -CA1 and GB_Mtb β -CA3, respectively. These descriptors effectively capture the hydrogen bonding potential of the molecule, which is indeed a crucial factor for binding to the active or allosteric sites in an enzyme to inhibit its biological function. Here, nHbint8 and nHbint6 refer to the number of hydrogen bond acceptors, not in a ring system but within the 8 bonds and 6 bonds of each other in a molecule [29]. This suggests that the active site of Mtb β -CA1 provides an architecture such that molecules with long-range hydrogen bonding interactions (8 bonds) are accommodated, and the Mtb β -CA3 accommodates molecules with medium-range interactions (6 bonds).

The top-ranking descriptor for the GB_Mtb β -Ca3 model was the ring count descriptor, i.e. **n5Ring**, which corresponds to the presence of 5-membered rings. This was found to be consistent with the important features of the mechanistic interpretation model, in which only the descriptor **n5 heteroring** was found to be positively correlated with the Mtb β -CA2 pKi, thus highlighting the importance of 5-membered ring systems, which are preferable heteroaromatic.

The top ten descriptors of the GB models had low correlation coefficients (Fig. 6c, f & 6i), suggesting that the set of diverse sets of independent descriptors selected by the model were not correlated with each other. This also improves the model's generalizability and interpretability, as the low intercorrelation reduces the risk of multicollinearity in the modeling. Furthermore, the incorporation of diverse sets of descriptors suggests that the models can better accommodate the complexity and variability that could be inherent in acquired datasets to make robust predictions [30,31]. The predicted activities of the designed molecules showed satisfactory performance (Table 7), with a threshold of pKi equivalent to or above 4 considered active and suitable for synthesis. The lower activity threshold was chosen due to the

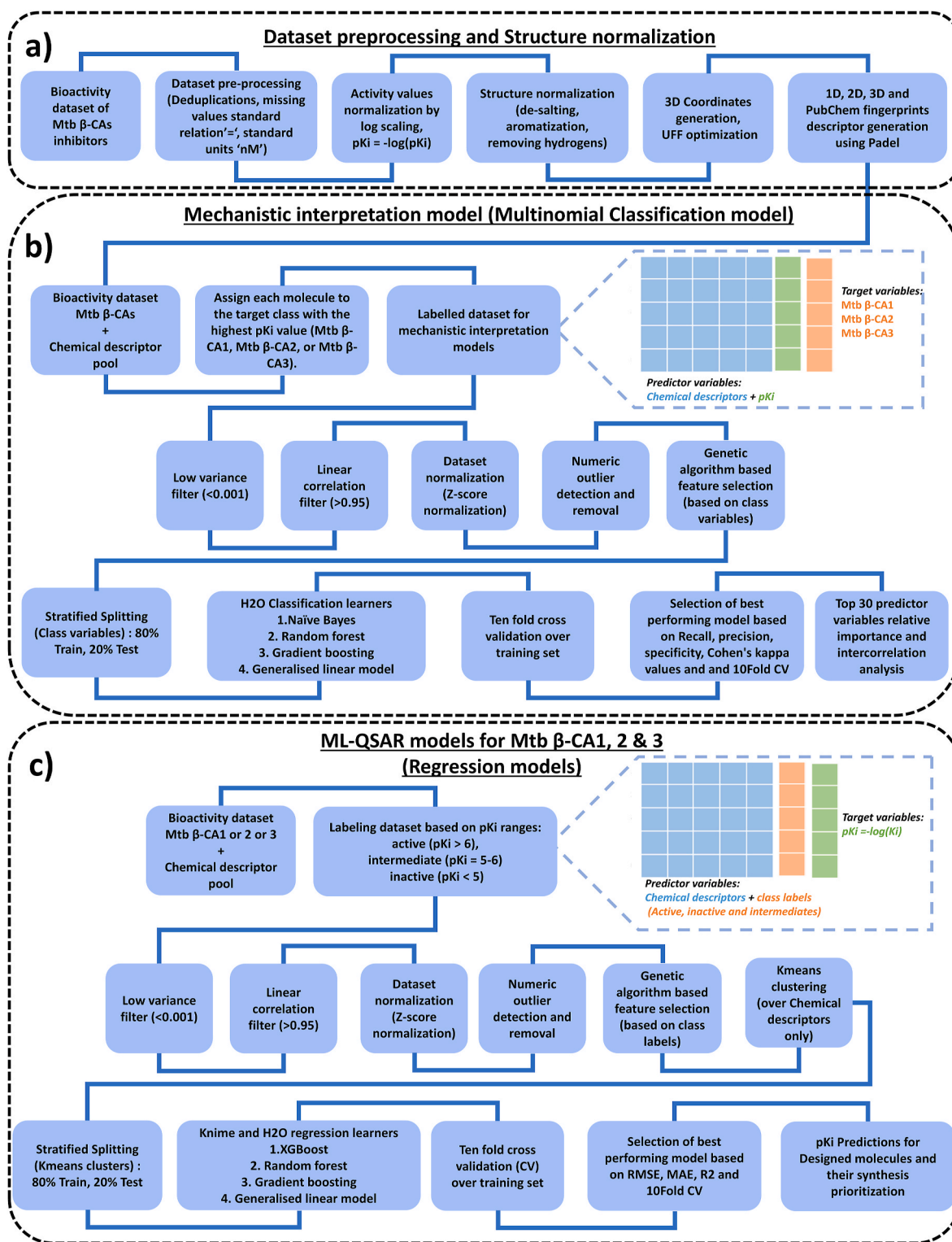


Fig. 4. Detailed workflow concerning a) dataset preprocessing and structure normalization, b) mechanistic interpretation, and c) ML-QSAR models.

Table 3

10-fold CVs of the Mechanistic interpretation model.

Model	Mean (Accuracy)	SD (Accuracy)
NB	0.58	0.11
RF	0.87	0.07
GB	0.88	0.10
GL	0.78	0.11

absence of DTC moieties in the training datasets, making it necessary to explore these structures' potential efficacy. Additionally, ring systems such as piperazine and hemipiperazine were novel for Mtb β -CAs chemical space exploration, warranting their inclusion despite their lower predicted activity. This approach aimed to expand the chemical diversity and explore new binding interactions within the Mtb β -CAs, which could lead to discovering potent inhibitors with unique mechanisms of action.

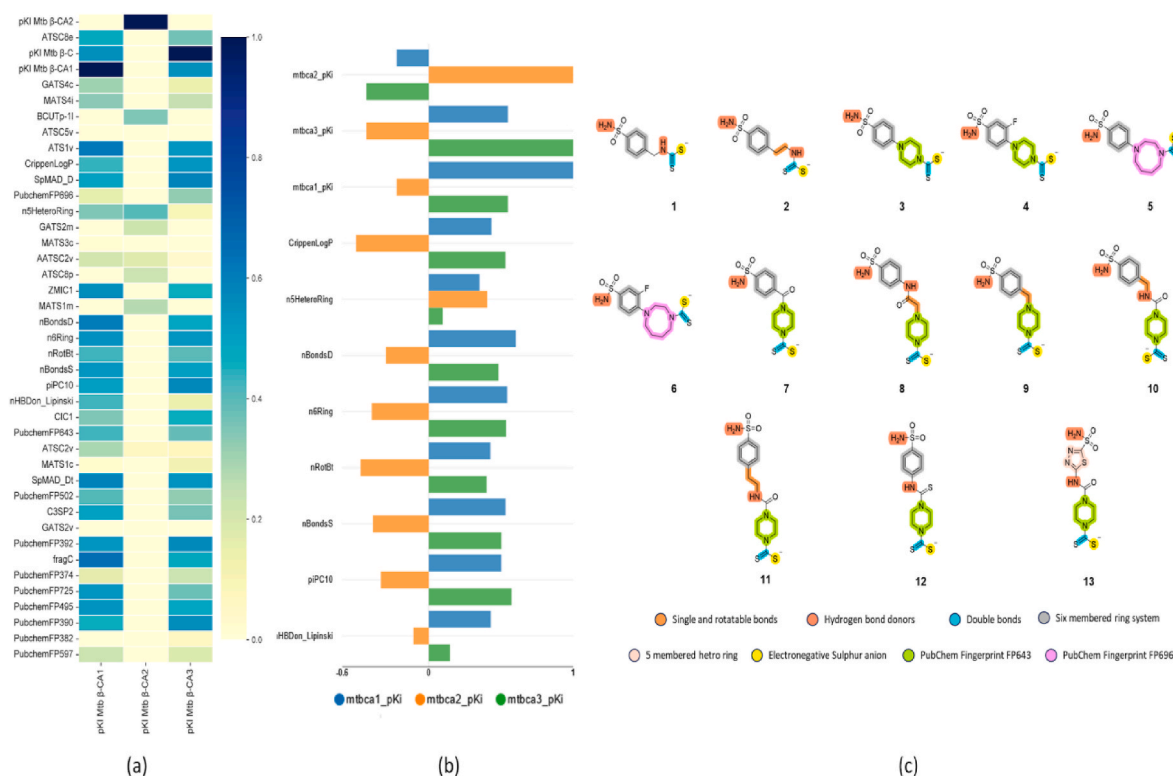


Fig. 5. Results of mechanistic interpretation modeling: a) Pearson correlation between Mtb β -CAs pKi and the top 30 important molecular descriptors visualized through the heatmap, b) bar plots indicating the correlation between Mtb β -CAs and chemical structure informant descriptors (count-based and PubChem fingerprints) c) Visualization of designed molecules incorporating different chemical motifs, color-coded based on their corresponding descriptors, is crucial for understanding their potential in Mtb β -CAs inhibition.

Table 4

Translation of mechanistic interpretations to molecular ideation.

Best Features	Description	Interpretations	Designed as
ATSC8e	Autocorrelation Electronegative descriptor	Electronegative motifs	Sulfur anion
n5HeteroRing	Number of 5- membered hetero rings	Five-membered, hetero ring	thiazole ring
nBondsD	Double bond structural counts	Double bonded structures	thione sulfur
n6Ring	Number of six- membered rings	Six-membered ring systems	Benzene (in benzene sulfonamide)
nRotBt	Number of rotatable bonds	Presence of flexible rotatable bonds	primary/ secondary aliphatic linkers
nBonds	Number of single bonds	Presence of single bonds	primary/ secondary aliphatic linkers
nHBDon_Lipinski	number of hydrogen bond donors	Presence of hydrogen bond donors	primary/ secondary amines in ZBG (Zinc Binding Groups) and linker
PubchemFP696	PubChem structural fingerprint representing C-C- C-C-C-C-C-C substructure	Substructure accounting for 8 membered system	Hemipiperazine rings
PubchemFP643	PubChem structural fingerprint representing [#1]- C-C-N-[#1]	two carbon atoms connected by single bonds followed by a nitrogen atom	Piperazine and Hemipiperazine rings

2.3. Applicability domains of the regression models

The applicability domains (ADs) of all the developed models were assessed using a principal component analysis (PCA) bounding box [32] on the generated molecular descriptors and fingerprints. The results were visualized in 2D and 3D to evaluate the coverage of the training sets, test sets, and designed molecules. As shown in Fig. 8, both the test and designed query compounds fell well within the chemical space covered by the training set. This suggests that the training set effectively encompasses all possible chemical motifs and structures, capturing the diversity of the chemical descriptors. Consequently, no molecule (test or designed query compound) lies outside the applicability domain of the models.

2.4. Synthetic chemistry

Herein, an appropriate benzenesulfonamide scaffold (accounting for six-membered rings and hydrogen bond donors) was modified based on a previously established mechanistic interpretation model by introducing a nucleophilic and reactive amine that could undergo nucleophilic addition to carbon disulfide (CS_2), yielding a dithiocarbamate group in which the Sulfur anion accounts for the electronegative molecular descriptor (ATSC8e) and the thione Sulfur double bond structural motif.

Various linkers were strategically designed to incorporate essential features such as single bonds, rotatable bonds, and hydrogen bond donor motifs and integrate reactive amine derivatives into the core benzenesulfonamide scaffold. Among these, urea, thiourea, and amide groups were explicitly considered for their potential to serve as effective linkers. Nevertheless, the absence of a specific linker between the different zinc-binding groups (ZBGs) was evaluated using different primary or secondary aliphatic groups. We primarily evaluated piperazine and

Table 5

Evaluation matrices of the tested ML algorithms for the regression QSAR model (MAE: mean absolute error, MSE: mean squared error, MSD: mean signed difference, MAPE: mean absolute percentage error).

Target	Models and sets	R ²	MAE	MSE	RMSE	MSD	MAPE
Mtb β -CA1	XGBoost_Train	0.85	0.34	0.18	0.43	0	0.06
	XGBoost_Test	0.7	0.44	0.31	0.56	0.22	0.08
	Random_Forest_Train	0.98	0.09	0.02	0.14	0.01	0.02
	Random_Forest_Test	0.93	0.18	0.07	0.27	0.05	0.03
	GB_Train	1	0.02	0	0.04	0	0
	GB_Test	0.97	0.08	0.03	0.18	0.01	0.01
	GL_Train	0.64	0.53	0.45	0.67	0.01	0.09
	GL_Test	0.58	0.54	0.46	0.68	0.12	0.09
Mtb β -CA2	XGBoost_Train	1	0.02	0	0.02	0	0
	XGBoost_Test	1	0.04	0	0.06	0.02	0.01
	RF_Train	0.99	0.08	0.01	0.12	0	0.01
	RF_Test	0.98	0.09	0.02	0.14	0.05	0.02
	GB_Train	1	0.02	0	0.05	0	0
	GB_Test	1	0.02	0	0.03	0.02	0
	GL_Train	0.81	0.3	0.2	0.45	0.03	0.06
	GL_Test	0.95	0.17	0.04	0.2	0.07	0.03
Mtb β -CA3	XGBoost_Train	0.88	0.26	0.13	0.36	0	0.04
	XGBoost_Test	0.42	0.47	0.42	0.65	0.19	0.08
	RF_Train	0.97	0.12	0.03	0.18	−0.01	0.02
	RF_Test	0.9	0.25	0.11	0.34	−0.03	0.04
	GB_Train	0.99	0.02	0.01	0.08	0	0
	GB_Test	0.97	0.11	0.04	0.19	0.03	0.02
	GL_Train	0.61	0.5	0.39	0.62	0.05	0.08
	GL_Test	0.59	0.53	0.45	0.67	−0.16	0.08

Table 6

Tenfold cross-validation results of the ML-QSAR models (MAE: mean absolute error, MSE: mean squared error, MSD: mean signed difference, MAPE: mean absolute percentage error).

Target	Models and sets	R ²	MAE	MSE	RMSE	MSD	MAPE
Mtb β -CA1	XGB_10CV	0.75	0.45	0.31	0.56	0	0.07
	RF_10CV	0.88	0.27	0.14	0.38	0	0.05
	GB_10CV	0.84	0.21	0.2	0.45	0	0.03
	GL_10CV	0.59	0.57	0.51	0.71	0.04	0.1
Mtb β -CA2	XGB_10CV	0.74	0.2	0.27	0.52	−0.15	0.03
	RF_10CV	0.85	0.27	0.15	0.39	0.01	0.05
	GB_10CV	0.92	0.15	0.08	0.28	0.02	0.03
	GL_10CV	0.64	0.41	0.38	0.61	0	0.07
Mtb β -CA3	XGB_10CV	0.59	0.46	0.45	0.67	−0.07	0.07
	RF_10CV	0.69	0.34	0.3	0.55	−0.07	0.05
	GB_10CV	0.79	0.22	0.2	0.45	−0.03	0.03
	GL_10CV	0.52	0.57	0.48	0.69	0	0.09

homopiperazine groups, identified through the PubChem fingerprints **PubChemFP643** and **PubChemFP696**, as key amines that connect ZBGs. These amines were chosen for their structural compatibility and functional properties, aiming to efficiently bridge the ZBGs. (Fig. 9). Also, the spectral plots (¹³C-NMR and ¹H NMR) for the synthesized compounds 1–13 are provided in [supplementary 2](#).

The first two products, 4-(aminomethyl)benzenesulfonamide hydrochloride (**14**) and 4-(aminoethyl)benzenesulfonamide (**15**), were obtained from commercially available aliphatic amines containing a benzenesulfonamide group. Both were treated in the presence of carbon disulfide and sodium hydroxide (5 M) in MeOH to produce the respective dithiocarbamate products **1** and **2** (Scheme 1).

All the other synthesized products were derivatized with a piperazine or homopiperazine group, which gave the best results considering yield and purity. First, both secondary amines were reacted in the presence of commercial 4-fluorobenzenesulfonamide (**16**) and 3,4-difluorobenzenesulfonamide (**17**) via a green chemical reaction named aromatic nucleophilic substitution. The latter was performed in H₂O at reflux to produce pure intermediates **18–21**. After that, the samples were treated with 5 M CS₂ and NaOH in MeOH at 0 °C to afford dithiocarbamates **3–6** (Scheme 2).

Compared to the next products, the previous compounds do not show any functionalized linkers between the benzenesulfonamide moiety and the secondary amine. Next, a coupling reaction was performed to connect the piperazine moiety to the benzenesulfonamide scaffold the with PyBOP as the coupling reagent and DIPEA in anhydrous DMF under an inert atmosphere, affording intermediate **23**. This reaction was performed using *t*-butyl piperazine-1-carboxylate and 4-sulfamoylbenzoic acid (**22**). Then, the Boc-protection was eliminated in acid media to obtain the proper chlorine salt (**24**), which was treated with 5 M CS₂ and NaOH in MeOH at 0 °C to produce dithiocarbamate **7** (Scheme 3). A different linker was designed starting from sulfanilamide (**25**), which was derivatized with chloroacetyl chloride to give the acetamide-based compound **26**. Then, a nucleophilic substitution was performed in the presence of *t*-butyl piperazine-1-carboxylate to synthesize intermediate **27**. Moreover, the same reaction was performed between freshly prepared 4-sulfamoylbenzyl bromide (**29**) and the piperazine derivative, yielding compound **30** in the presence of K₂CO₃ in anhydrous DMF. After piperazine deprotection in acid media (**28,31**), dithiocarbamates were obtained following the same method, giving products **8** and **9** (Scheme 3).

Following the intent to expand the structure-activity relationships, even urea and thiourea linkers were used to connect the benzenesulfonamide scaffold and the piperazine group. The urea derivatives were obtained starting from proper commercial aliphatic amines **14** and **15**, which were treated in the presence of phenylchloro formate and K₂CO₃ in acetone, predisposing the NH₂ group for the next nucleophilic substitution (**32,33**). The latter was performed with *t*-butyl piperazine-1-carboxylate and DIPEA in anhydrous ACN to give intermediates **34** and **35**. Nevertheless, the thiourea group was synthesized from freshly prepared 4-isothiocyanatobenzenesulfonamide (**38**) in the presence of *t*-butyl piperazine-1-carboxylate to give compound **39**. All these intermediates were deprotected in the presence of 3 M MeOH/HCl or 4 M dioxane/HCl to yield **36**, **37**, and **40**, respectively. Finally, nucleophilic addition was performed following the previously described conditions to obtain **10–12** (Scheme 4).

Finally, the last product was synthesized starting from the clinically available acetazolamide (**41**), which was deacetylated in the presence of HCl in EtOH at reflux, leading to the subsequent nucleophilic substitution of the obtained intermediate **42** with phenylchloro formate in anhydrous THF at reflux. This reaction afforded carbamate **43** in the

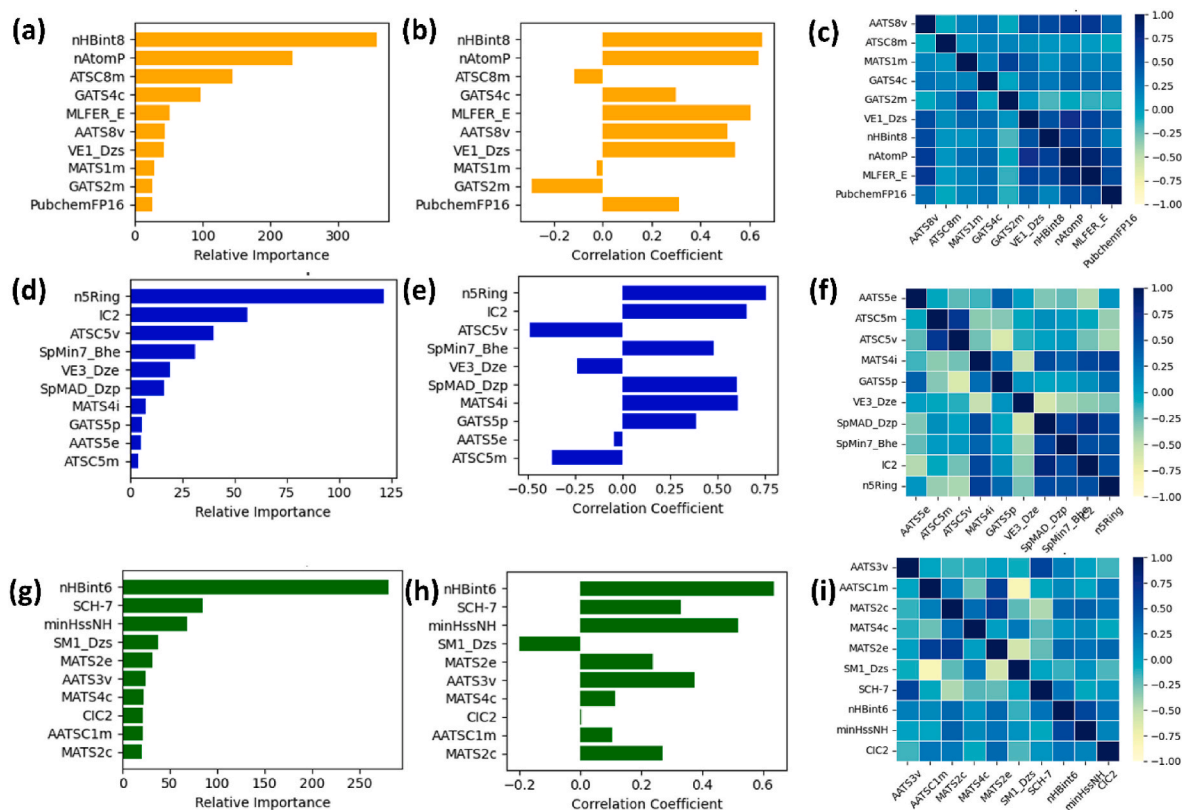


Fig. 6. The top 10 variables by their relative importance, correlation with biological endpoints, and their intercorrelations are shown for the GB QSAR models a, b, c) Mtb β -CA1, d, e, f) Mtb β -CA2 and g, h, i) Mtb β -CA3.

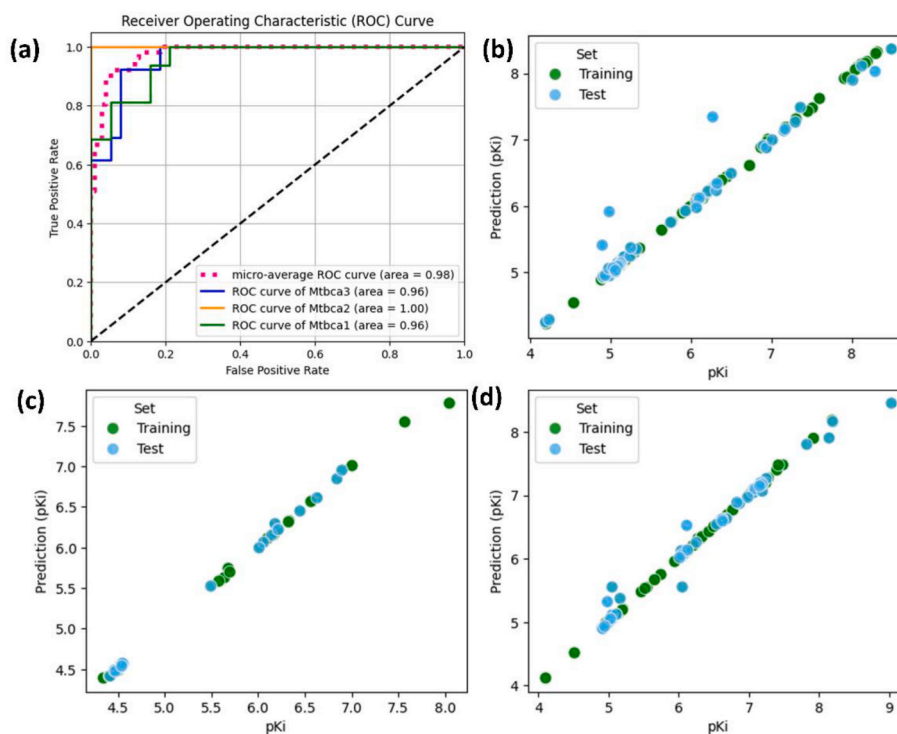
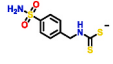
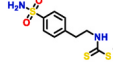
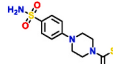
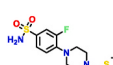
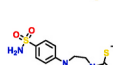
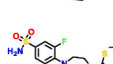
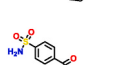
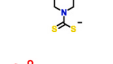
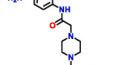
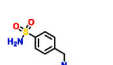
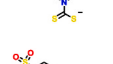
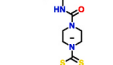
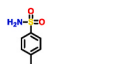
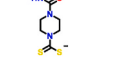


Fig. 7. Performance of the developed models. a) ROC plots for the mechanistic interpretation models for Mtb β -CA1-3 classification (random forest) in a one-versus-all manner and Predicted pKi vs experimental pKi values of ML-QSAR models (gradient boosting) for a) Mtb β -CA1, b) Mtb β -CA2 and c) Mtb β -CA3.

Table 7
Predicted biological activity of the prediction set.

Compounds	2D structures	Predicted pKi (pKi = -log (KI))			Synthesis status (pKi ≥ 5)
		Mtb β-CA1	Mtb β-CA2	Mtb β-CA3	
1		5.88	4.58	5.94	✓
2		6.13	5.21	5.68	✓
3		5.58	4.83	5.93	✓
4		5.36	5.18	6.17	✓
5		5.83	5.88	4.73	✓
6		5.61	5.8	5.45	✓
7		5.81	5.57	5.59	✓
8		6.4	5.77	6.24	✓
9		5.76	5.81	5.9	✓
10		6.29	5.69	5.95	✓
11		6.08	5.57	6.02	✓
12		7.54	6.06	6.85	✓
13		6.03	6.16	6.4	✓
AAZ		6.32	7.78	7.92	

presence of *t*-butyl piperazine-1-carboxylate to give compound **44**, which was deprotected in an acidic environment (**45**). Then, nucleophilic addition in the presence of CS₂ and 5 M NaOH in MeOH was performed to give dithiocarbamate **13** (Scheme 5).

2.5. Biological evaluation

Derivatives **1–13** were tested for their inhibitory action against the target β-CA Mtb 1, 2, and 3 and the off-target Human carbonic anhydrase (hCA) isoforms I and II. Acetazolamide (**AAZ**) was used as a standard CA inhibitor. CA I and II are the main off-target isoforms for our desired therapeutic application. The following structure-activity relationships (SAR) can be gathered from the inhibition data reported in Table 8. Cytosolic CA I was inhibited by benzenesulfonamide derivatives **1–13**, with inhibition constants (K_i) ranging from the low-medium nanomolar to the low micromolar range (K_i: 97.5–7322 nM). This isoform was the least inhibited isoform in the kinetic study, probably due to its peculiar active site architecture and amino acid composition in comparison to most other CA isoforms. Interestingly, the acetazolamide analog **13** reached one of the best inhibitory activities, showing K_i values in the medium nanomolar range of 522.8 nM.

On the other hand, the benzenesulfonamide scaffold gave a wide range of values, probably due to the difference in the linker between the two ZBGs. In fact, the absence of a specific functional linker generated the worst effects on hCA I, **3** (K_i: 2300 nM), **4** (K_i: 3663 nM), **5** (K_i: 3437 nM), and **6** (K_i: 7322 nM). The spacer enhancement between the piperazine group and the benzenesulfonamide scaffold shifted the K_i values to the medium-high nanomolar range, as demonstrated by compounds **8** (K_i: 450.9 nM), **9** (K_i: 598.0 nM), **10** (K_i: 693.3 nM), **11** (K_i: 830.9 nM) and **12** (K_i: 628.5 nM). For this isoform, the differences in the functional linker did not result in a wide gap in the K_i values, except for product **7**, which reached a K_i value in the low micromolar range of 4090 nM.

Cytosolic hCA II was effectively inhibited by most benzenesulfonamide derivatives, with K_i values ranging from the sub-nanomolar to medium-high nanomolar range (K_i: 7.1–797.5 nM). Product **9** stood out as the best inhibitor for this isoform, reaching a K_i value in the low nanomolar range of 7.1 nM, approximately 10-fold more potent than analog **7** (K_i: 76.5 nM), which has a ketone group between the ZBG and the piperazine moiety. Urea derivative **10** had a high affinity for this human isoform, with a K_i in the low nanomolar range of 10.3 nM. Moreover, compared to its superior analog **11**, the latter stood out as a weak inhibitor, showing a K_i value of 93.1 nM and 9-fold less potency. On the other hand, products **3** (K_i: 350.1 nM), **4** (K_i: 797.5 nM), **5** (K_i: 755.9 nM), and **6** (K_i: 399.7 nM) were the weakest inhibitors of hCA II, reaching K_i values in the medium-high nanomolar range. Finally, derivative **13**, a thiadiazole-like compound, demonstrated less affinity than standard **AAZ**, ranging from 12.0 nM to 84.6 nM.

Mtb β-CA1 was one of the most efficiently inhibited isoforms by all the compounds, with K_i values in the range of 19.9–519.4 nM. In contrast to hCA I and II, products **3**, **7**, **8**, and **9** were the weakest inhibitors of β-CA, with K_i values of 193.5 nM, 519.5 nM, 285.3 nM, and 217.1 nM, respectively. Interestingly, a pivotal increase in the K_i value was observed upon moving from compound **3** to analog **4**, as indicated by the presence of a fluorine atom on the phenyl ring and the presence of a homopiperazine group instead of piperazine at positions 5 and 6. These reached K_i values in the low nanomolar range of 50.9 nM, 19.9 nM, and 66.9 nM, respectively. All these compounds had more potent activity on β-CA1 than did standard **AAZ** (480.0 nM). Moreover, thiadiazole-based **13** showed a K_i in the low nanomolar range of 85.7 nM, approximately 5-fold greater than that of the standard.

Mtb β-CA2 is the most inhibited isoform from standard **AAZ**, showing a K_i value of 9.8 nM, while all the CAI (CA inhibitor) derivatives reached K_i values in the low-medium nanomolar range, between 6.9 nM and 506.1 nM. The most potent compound was number **2**, with an inhibition constant in the same range as that of the standard **AAZ** (6.9 nM). On the

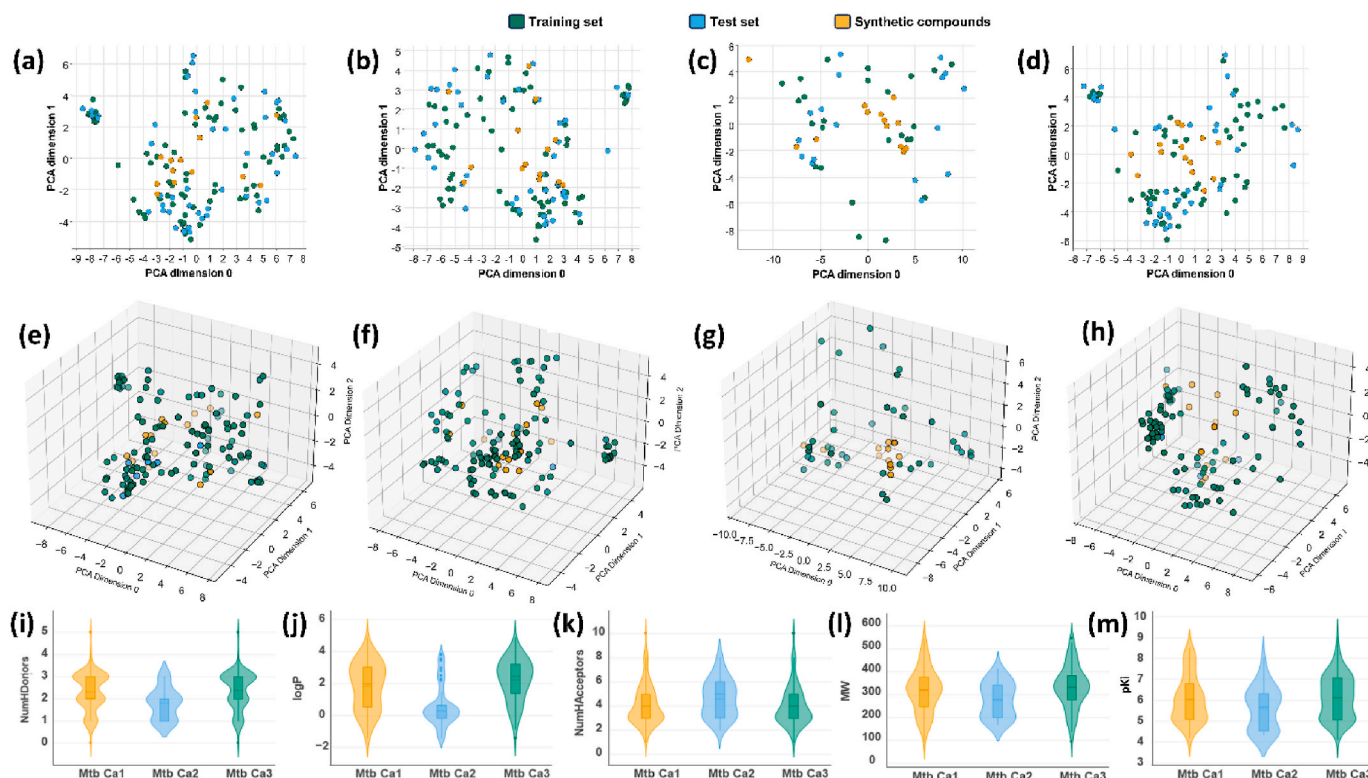


Fig. 8. Applicability domains of the mechanistic interpretation models (a,e) and ML-QSAR models for Mtb β -CA1 (b,f), Mtb β -CA2 (c,g), and Mtb β -CA3 (d,h). Dataset exploration via violin plots for the Mtb β -CAs datasets: (i) Number of hydrogen bond donors, (j) LogP, (k) Number of hydrogen acceptors, (l) Molecular weight, and (m) experimentally determined pKi.

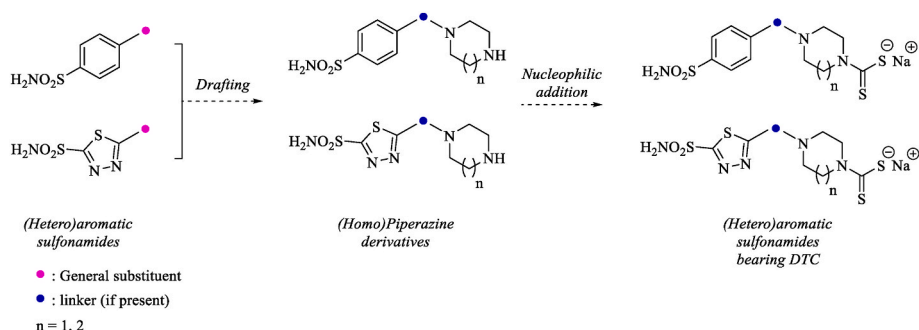
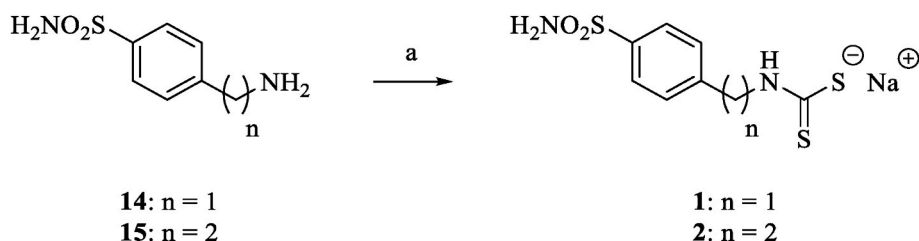


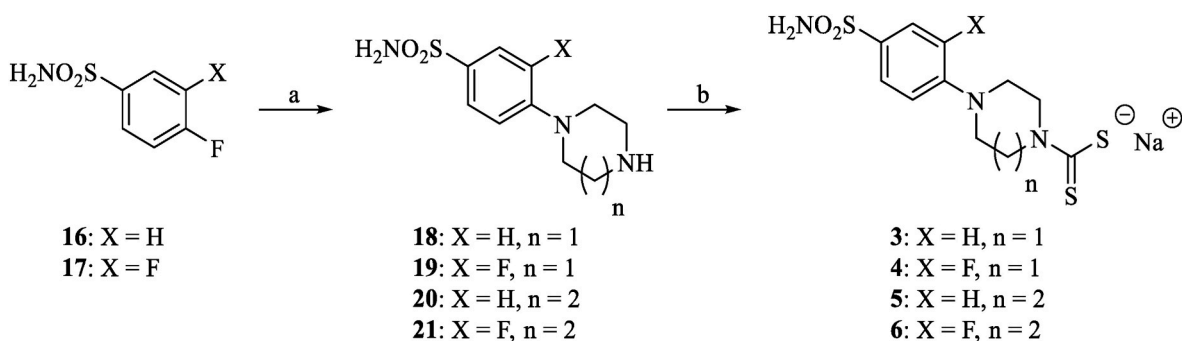
Fig. 9. Proposed drug design strategy to synthesize innovative antitubercular agents based on mechanistic interpretation models and ML-QSAR prioritization.



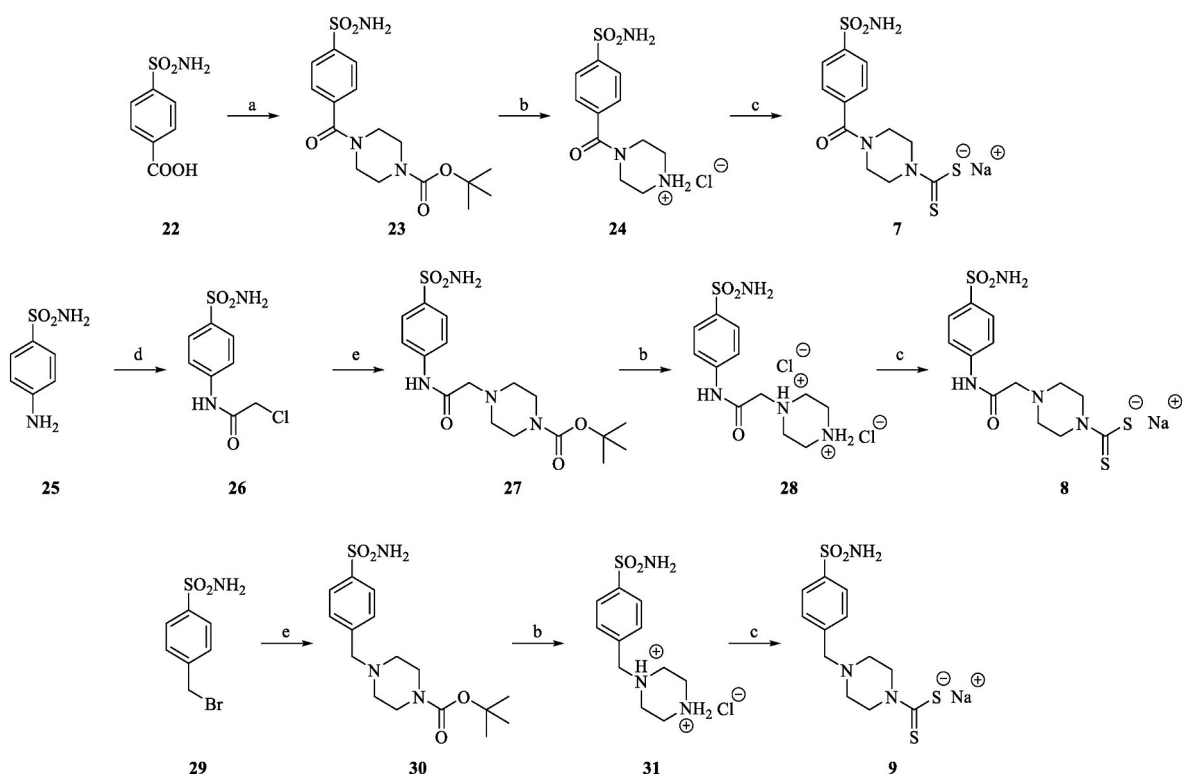
Scheme 1. Synthetic approach starting from commercial aliphatic amine intermediates 14 and 15 to achieve proper DTCs 1 and 2. Reagents and conditions: a) CS₂, 5 M NaOH, MeOH, 0 °C to RT, 1 h.

other hand, CA inhibitors **3** (K_i : 275.8 nM), **9** (K_i : 170.0 nM), **10** (K_i : 506.1 nM), **12** (K_i : 418.2 nM), and **13** (K_i : 321.1 nM) strongly reduced the affinity of this isoform, resulting in K_i values in the medium nanomolar range. In particular, compounds **10** (K_i : 506.1 nM) and **12** (K_i : 418.2 nM) were the worst inhibitors of the cited enzyme. However,

moving from inhibitor **9**, which has a methylene linker, to compound **7**, which has a ketone linker, improved the potency from 170.0 nM to 43.2 nM. Similarly, increasing the distance between the benzenesulfonamide scaffold and piperazine increased the affinity for Mtb β -CA2, as reported for ZBGs **10** and **11**, which had a K_i of 80.2 nM.



Scheme 2. Synthetic approach to achieve proper DTCs 3–6. Reagents and conditions: a) Piperazine or homopiperazine, H₂O, reflux, o.n.; b) CS₂, NaOH 5 M, MeOH, 0 °C to RT, 1 h.



Scheme 3. Synthetic approach to achieve proper DTCs 7–9. Reagents and conditions: a) t-butyl piperazine-1-carboxylate, PyBOP, DIPEA, dry DMF, 0 °C to RT, 0 °C; b) MeOH/HCl 3 M, RT, 0 °C; c) CS₂, NaOH 5 M, MeOH, 0 °C to RT, 1 h; d) chloroacetyl chloride, K₂CO₃.

Mtb β -CA3 was one of the most efficiently inhibited isoforms by all the compounds, with K_i values in the range of 20.5–371.3 nM. In contrast to isoform 2, which has been previously discussed, compound 7 had one of the worst K_i values, 299.6 nM, in the medium nanomolar range. On the other hand, the affinity of product 3 for all the Mtb isoforms was low, with a K_i value of 371.3 nM, but its analogs, such as 4, 5, and 6, had the best K_i values of 26.8 nM, 33.2 nM, and 20.5 nM, respectively. In this scenario, the presence of a fluorine atom on the benzenesulfonamide scaffold does not seem to affect the inhibition constant (See Table 8). Similar observations could be made for compounds 10 (K_i : 59.8 nM) and 11 (K_i : 60.6 nM), where the distance between the ZBG and piperazine did not affect the K_i .

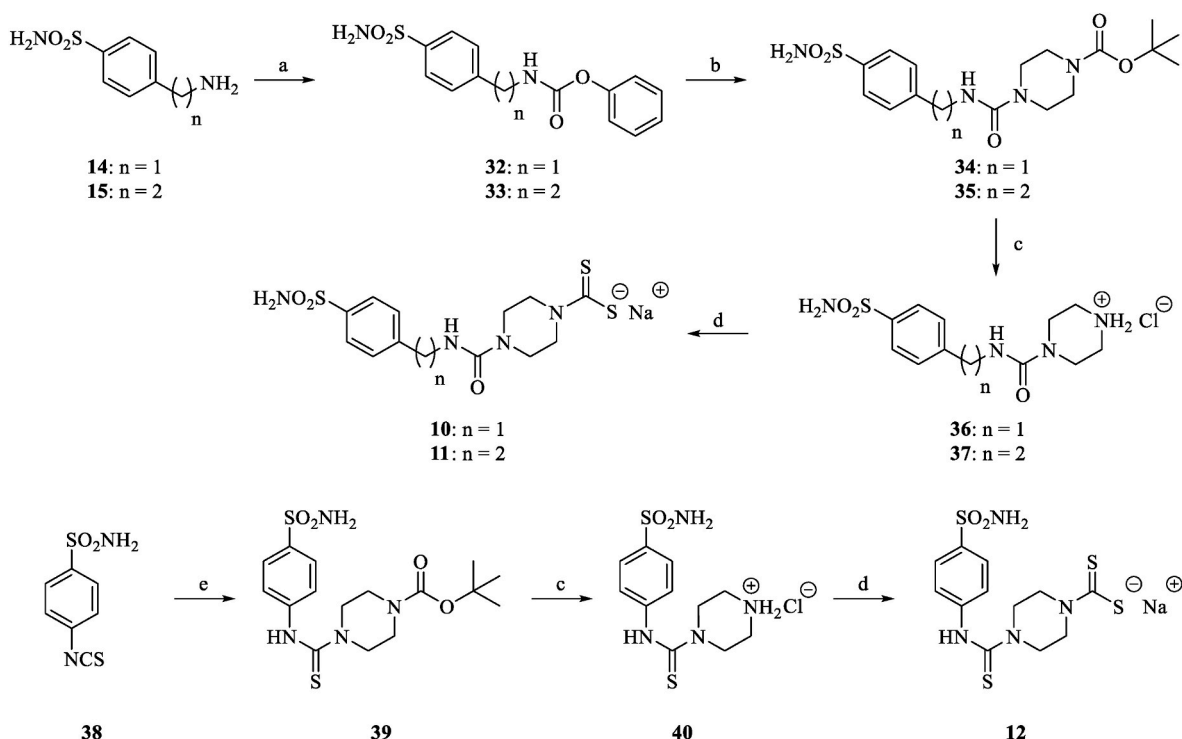
2.6. ADMET and molecular docking analysis

For the molecules to be effective and safe, they must exhibit favorable pharmacokinetic properties and optimal toxicological profiles. To ensure this, the synthesized compounds were thoroughly evaluated for

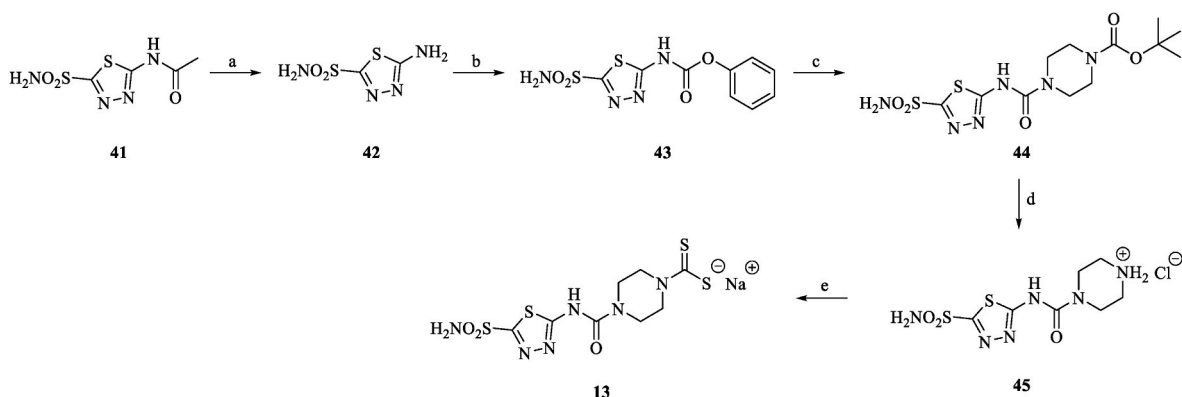
their ADME (Absorption, Distribution, Metabolism, and Excretion) characteristics and toxicity profiles (Supplementary 1, Tables S4 and S5 for ADME and Toxicity, respectively). This evaluation is crucial to confirm their pharmacological efficacy while minimizing adverse effects on recipients.

Most of the compounds demonstrated high gastrointestinal (GI) absorption, except for molecules 12, 13, and AAZ, which had low GI absorption. None of the compounds showed blood-brain barrier (BBB) permeation, suggesting they are unlikely to cause central nervous system related adverse effects. Importantly, none of the compounds were P-glycoprotein (P-gp) substrates, meaning P-gp would not efflux them out of cells. This characteristic is advantageous, allowing the compounds to effectively penetrate and remain within host cells, achieving sufficient concentrations to act against pathogens.

Moreover, many drug-drug interactions result from the inhibition or induction of P-gp. Non-P-gp substrates are less likely to be involved in such interactions, facilitating their co-administration with other medications. This not only improves therapeutic outcomes but also increases



Scheme 4. Synthetic approach to achieve proper DTCs 10–12. Reagents and conditions: a) phenylchloro formate, K_2CO_3 , acetone, 0 °C to RT, 2 h; b) t-butyl piperazine-1-carboxylate, DIPEA, dry ACN, reflux, o.n.; c) MeOH/HCl 3 M, RT, o.n. or dioxane/HCl 4 M, RT, o.n.; d) CS_2 , NaOH 5 M, MeOH, 0 °C to RT, 1 h; e) t-butyl piperazine-1-carboxylate, DIPEA, dry ACN, RT, o.n.



Scheme 5. Synthetic approach to achieve proper DTC 13. Reagents and conditions: a) 12 M HCl, EtOH, reflux, o.n.; b) phenylchloro formate, dry THF, 0 °C to reflux, 2 h; c) t-butyl piperazine-1-carboxylate, DIPEA, dry ACN, reflux, o.n.; d) 3 M MeOH/HCl, RT, o.n.

the safety of medications [33]. By avoiding efflux via P-gp, non-P-gp substrates may reduce the risk of drug resistance. Efflux pumps like P-gp are a mechanism by which bacteria and host cells develop resistance to drugs, so circumventing this pathway can be beneficial in maintaining drug efficacy [34].

The metabolism profiling of the synthesized compounds reveals varying inhibitory effects on cytochrome P450 (CYP) enzymes, which are crucial for drug metabolism. Compounds 1 and 2 inhibit CYP1A2 but do not inhibit CYP2C19, CYP2C9, CYP2D6, or CYP3A4, suggesting they may interfere with drugs metabolized by CYP1A2, such as isoniazid, which is a first-line antitubercular drug [35,36]. Compounds 3, 4, 5, 6, and 9 inhibit CYP2C19, indicating potential interactions with drugs metabolized by this enzyme, such as rifampicin, which is also metabolized by CYP2C19 and CYP3A4 [37]. Compound 12 exhibits a broader inhibition profile, affecting both CYP1A2 and CYP2C19, potentially interfering with the metabolism of isoniazid and rifampicin.

Compounds 7, 8, 10, 11, 13, and AAZ do not inhibit any of the tested CYP enzymes, suggesting a lower likelihood of metabolic interactions and potentially fewer side effects related to CYP-mediated drug metabolism. This comprehensive profiling helps identify compounds with favorable metabolic properties and guides further development to minimize adverse drug interactions, particularly with essential antitubercular drugs like isoniazid and rifampicin.

The toxicological analysis of the synthesized compounds indicates that the majority exhibit a favorable safety profile. Compounds 1 through 7, 9, 10, 11, and 12 show no mutagenic, tumorigenic, reproductive, or irritant effects, suggesting minimal toxicity concerns. Compound 8, while not mutagenic, tumorigenic, or a reproductive effector, is identified as a high-level irritant, which may require caution in handling. Compound 13 does not exhibit mutagenic, tumorigenic, or irritant properties but is noted for its high potential as a reproductive effector. Compound AAZ presents a higher risk profile, being non-

Table 8Inhibition data of hCA I and II and bacterial Mtb β -CA1, Mtb β -CA2, and Mtb β -CA3 using AAZ as a reference drug by a Stopped-Flow technique [22].

Code	K _i (nM) ^a			pK _i = -log ₁₀ (K _i)				
	hCA I	hCA II	Mtb β -CA 1	Mtb β -CA 2	Mtb β -CA 3	Mtb β -CA 1	Mtb β -CA 2	Mtb β -CA 3
1	733.4	55.2	65.7	31.2	106.6	7.18	7.51	6.97
2	97.5	48.1	55.3	6.9	43.3	7.26	8.16	7.36
3	2300	350.1	193.5	275.8	371.3	6.71	6.56	6.43
4	3663	797.5	50.9	71.6	26.8	7.29	7.15	7.57
5	3437	755.9	19.9	76.5	33.2	7.7	7.12	7.48
6	7322	399.7	66.9	92.0	20.5	7.17	7.04	7.69
7	4090	76.5	519.4	43.2	299.6	6.28	7.36	6.52
8	450.9	59.5	285.3	109.8	171.9	6.54	6.96	6.76
9	598.0	7.1	217.1	170.0	177.0	6.66	6.77	6.75
10	693.3	10.3	100.7	506.1	59.8	7	6.3	7.22
11	830.9	93.1	28.1	80.2	60.6	7.55	7.1	7.22
12	628.5	10.6	72.6	418.2	68.6	7.14	6.38	7.16
13	522.8	84.6	85.7	321.1	90.3	7.07	6.49	7.04
AAZ	250.0	12.0	480.0	9.8	104.0	6.32	8.01	6.98

^a The means from 3 different assays determined by a stopped-flow technique (errors were in the range of ± 5 –10 % of the reported values).

mutagenic and non-irritant but showing significant tumorigenic and reproductive effects. This comprehensive toxicological profiling aids in identifying compounds with acceptable safety profiles for further development and potential therapeutic use.

Molecular docking analysis was conducted to validate the experimentally determined pK_i using structure-based methodologies. The synthesized compounds were docked into the active site cavities of the targets Mtb β -CA 1, 2, and 3. For Mtb β -CA 1 and Mtb β -CA 2, the protein structures were obtained from the Protein Data Bank (PDB), while the structure for Mtb β -CA 3 was obtained through AlphaFold predictions. The modeled protein for Mtb β -CA 3 was evaluated for its structural quality using expected positional error estimates, Ramachandran plot analysis, and normalized QMEAN scores. The docking cavity used for Mtb β -CA 3 encompassed residues 600 to 765, and this region showed almost zero expected positional error (illustrated in [Supplementary 1, Fig. S8b](#)). Additionally, the modeled protein had 95.01 % of its residues in favored Ramachandran regions and 1.57 % outliers. The outliers were visualized and found to be far from the zinc-binding residues (See [Supplementary 1, Fig. S13](#)), suggesting that these artifacts might not affect the docking outcomes.

The modeled protein also demonstrated a MolProbity score of 1.77, indicating a structure with a resolution of approximately 1.77 Å, which is suitable for structure-based drug design (SBDD). Furthermore, when compared with the non-redundant set of PDB structures based on the normalized QMEAN score, the structure was found to fall within the acceptable range, with a Z-score of less than 1. This indicates that the structure resembles the geometry of high-quality protein structures deposited in PDB. In summary, a high-quality protein structure should have at least 90–95 % of its residues in favored Ramachandran regions and less than 1–2 % outliers, both of which are met by the modeled protein [38].

The docking performance was then correlated with the experimental pK_i values for each respective target ([Supplementary 1, Figs. S9d, S9e, S9f](#)). The Pearson correlation coefficients between the docking scores and the experimental K_i values were as follows: for Mtb β -CA 1, the correlation coefficient was $r = 0.52$, indicating a moderate positive correlation; for Mtb β -CA2, the correlation coefficient was $r = 0.47$, also showing a moderate positive correlation; and for Mtb β -CA3, the correlation coefficient was $r = 0.03$, indicating a very weak correlation. This suggests that the docking scores for Mtb β -CA 3 are not predictive of the experimental K_i values, likely due to differences in structural accuracy or the inherent variability in the AlphaFold prediction model compared to experimentally derived structures. Limitations such as variability in prediction accuracy, difficulties with disordered regions, and the lack of consideration for protein conformational flexibility further undermine the reliability of AlphaFold predictions in structure-based drug design [39,40]. These results suggest that the docking

scores are more predictive of the experimental K_i for Mtb β -CA 1 and Mtb β -CA 2 compared to Mtb β -CA3.

To deepen the understanding of the binding modes of compounds 2, 4, and 5 with the Mtb β -Cas, molecular interactions were plotted and analyzed. The 3D binding modes, illustrated in [Fig. 10](#) on the left side, suggest that compounds 2 and AAZ, as well as compounds 4 and 5, have similar binding poses. This similarity in binding poses is likely due to their structural similarities. Compound 2 and AAZ both contain only one ring system, whereas compounds 4 and 5 have two ring systems. The presence of two-ring systems in compounds 4 and 5 might affect their ability to accommodate into the binding pocket as AAZ does, thus having a different binding mode.

Compounds 2, 4, and 5 exhibited significant binding interactions with the zinc ion and the critical zinc-binding residues [41], specifically two cysteine residues and a histidine, across all three Mtb β -CAs I-III. This interaction effectively inhibits the catalytic activity of these metalloenzymes. Notably, compounds 4 and 5, which feature extended ring systems, displayed enhanced hydrophobic interactions compared to compound 2 and the known inhibitor AAZ. These hydrophobic interactions contribute to the stabilization of their binding poses once anchored into the catalytic triad, thereby reinforcing their inhibitory effects on the enzymatic activity of Mtb β -CAs. This characteristic highlights the importance of structural modifications in optimizing the binding affinity and efficacy of potential inhibitors targeting these enzymes.

2.7. Systems biology approaches to understand the probable mechanism of action

To ensure a coherent elucidation of the potential mechanisms of action, we exclusively conducted a systems analysis on five synthetic compounds, 12, 8, 13, 10, and 9, alongside the control standard AAZ (acetazolamide) (described in section 5.4.9, supporting data in [supplementary 1, Table S4](#)). These molecules were chosen based on their robust prediction accuracy within established mechanistic interpretation models and ML-QSAR models. The selection of these specific compounds was deliberate, aimed at producing consistent mechanisms of action that are grounded in both cheminformatics and systems biology methodologies. Multiple target prediction servers were utilized to collect a diverse pool of target predictions originating from different methodologies and rationales, and these ensemble approaches were found to be more effective in identifying probable real targets [42]. A total of 236 deduplicated target gene pools were collected for the synthesized compounds. The predicted target genes were plotted as a diagram against the disease gene pool of TB (19,994 genes), and 181 genes were found at the intersection ([Fig. 11 a](#)). These intersecting genes were used as a starting point to generate a complex PPI network with the aid

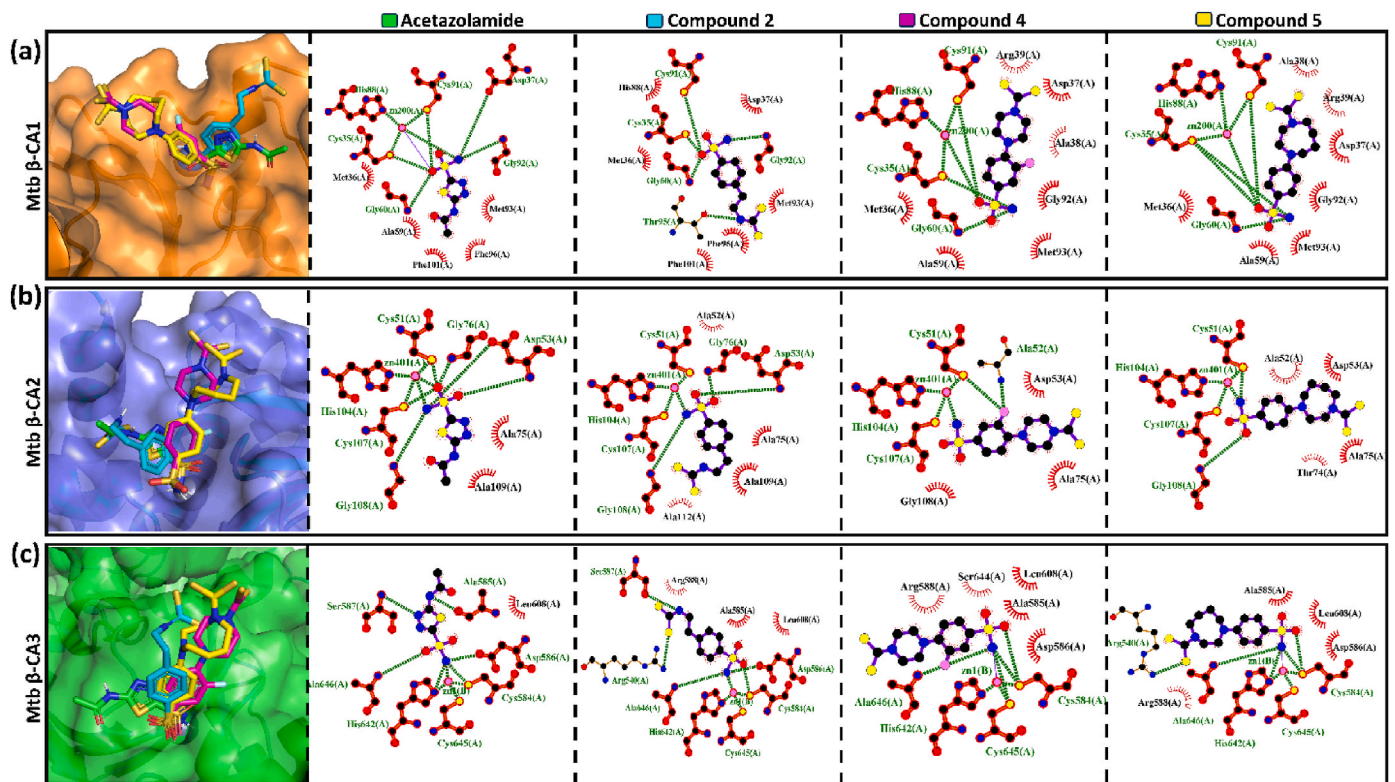


Fig. 10. Molecular docking analysis of the compounds AAZ, 2, 4, and 5 and for the target a) Mtb β -CA1, b) Mtb β -CA2, and c) Mtb β -CA3. The leftmost panel indicates the 3D binding poses of the ligands, and their color codes are as follows: Acetazolamide (AAZ): Green, Compound2: Blue, Compound4: Purple, and Compound5: Yellow.

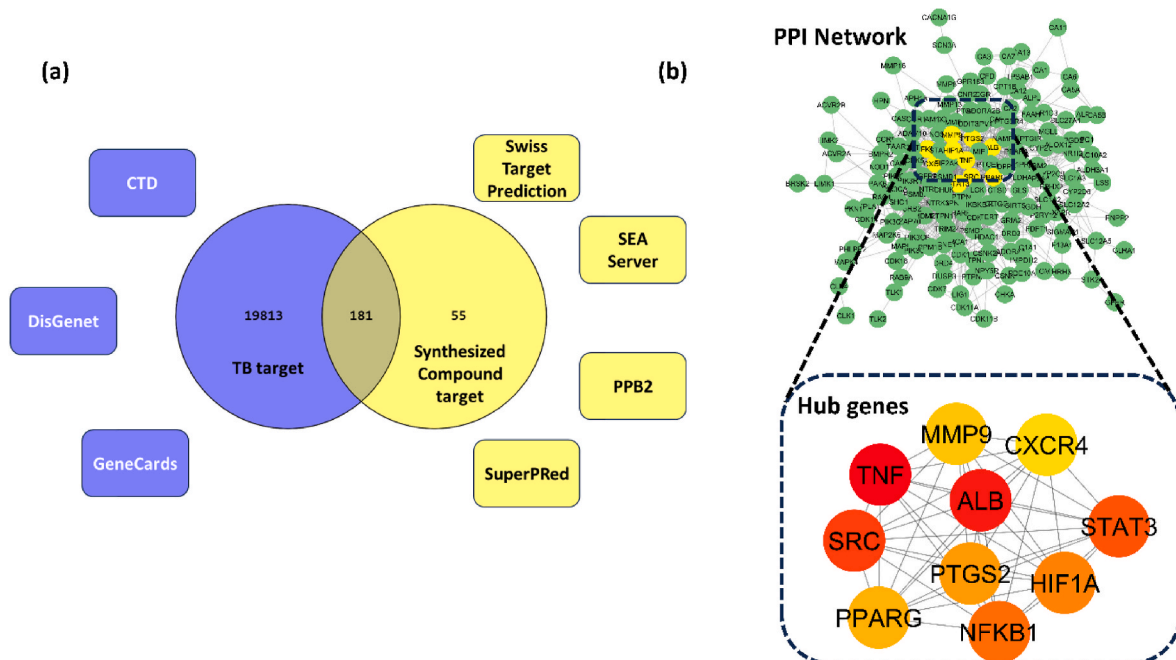


Fig. 11. Depicts the results of the systems biology analysis: a) intersecting genes between TB-related genes and synthesized compound targets, b) extraction of hub genes from the high-confidence PPI network.

of String DB. Additionally, interconnections with a confidence interval of 0.9 were used in this study. Afterward, ten hub genes (degree method) were extracted from this filtered PPI network, which resulted in the identification of MMP, CXCR4, TNF, ALB, STAT3, SRC, PTGS2, HIF1A, NFKB1, and PPARG hub genes (Table 9 shows topological matrices of

hubgenes and illustration of hubgenes extraction in Fig. 11b). The hub genes play a crucial role in the PPI network; where they act as key regulators within a defined PPI network, and they also serve as potential drug targets for therapeutic interventions [43].

Table 9
Topological metrics of the extracted hub genes.

Hub genes	Degree	EPC	BottleNeck	EcCentricity	Closeness	Betweenness
TNF	72	21.009	49	0.24704	116.9167	3655.757
ALB	69	19.471	17	0.24704	114.25	5415.203
SRC	64	20.401	27	0.24704	112.3333	3324.7
STAT3	57	20.196	1	0.19763	108.5333	1286.494
HIF1A	50	18.683	8	0.19763	104.2	1144.957
NFKB1	50	19.208	1	0.24704	104.25	774.5789
PTGS2	49	17.753	12	0.24704	103.5	1506.332
PPARG	47	17.48	5	0.24704	103.0833	1506.907
MMP9	43	18.423	1	0.19763	99.78333	604.2381
CXCR4	35	15.4	6	0.19763	96.11667	428.6465

2.7.1. Gene ontologies

The hub genes were annotated based on their biological processes (BP), cellular compartment (CC), and molecular function (MF) via gene ontology to understand hub genes based on their annotated scientific vocabulary. The gene ontology results are described in Fig. 12 in the form of gene enrichment bubble plots. The 5 biological processes ontologies with the highest enrichment and greatest counts were negative regulation of gene expression, positive regulation of *pir*-mRNA transcription from the RNA polymerase II promoter, positive regulation of apoptotic process, cellular response to hypoxia, and negative regulation of *pir*-mRNA transcription from the RNA polymerase II promoter. The CC ontologies included RNA polymerase II transcription factor complex, cytoplasm, macromolecular complex, chromatin, and caveola. The MF ontologies were mainly associated with DNA binding, small molecule binding, and enzyme binding, whereas many ontologies suggest DNA binding is related to RNA polymerase II transcription.

The gene ontologies indicated that hub genes are involved in gene expression via the regulation of *pir*-mRNA transcription from the RNA polymerase II promoter and localized the RNA polymerase II transcription factor complex, and the functions of these genes are mainly related to DNA binding.

The BP, CC, and MF gene ontologies indicated that the hub genes are involved in gene expression, particularly through the regulation of *pir*-mRNA transcription from the RNA polymerase II promoter. This is further supported by MF ontologies that highlight binding events such as DNA binding, small molecule binding, and enzyme binding. The BP terms positive regulation of the apoptotic process, positive regulation of cytokine production involved in the inflammatory response, and response to xenobiotic stimulus were all correlated with immune-related

processes. Additionally, the CC hits included the RNA polymerase II transcription factor complex, which is related to gene expression and its regulation, and caveolae, which are associated with pathogen invasion, control of cellular apoptosis, and inflammatory host responses [44].

2.7.2. KEGG pathway analysis

After analyzing the gene ontologies that indicated that the hub genes were related to immunity and inflammation-related pathways, we performed KEGG pathway analysis to better understand their involvement in biological pathways to obtain a larger landscape. KEGG hits not related to TB or infections were excluded, and only those hits that were statistically significant (p value < 0.05) were used for analysis and network construction. According to the KEGG pathway analysis (bubble enrichment plots in Fig. 13), the hits were related to many signaling pathways, such as the IL-17, C-type lectin receptor, TNF, chemokine, prolactin, and NF- κ B pathways. Additionally, hits related to disease-specific conditions included toxoplasmosis, efferocytosis, tuberculosis, pathogenic *Escherichia coli* infection, and COVID-19. From the gene ontology and KEGG pathway analyses, it was concluded that the synthesized molecules not only target *Mtb* β -Cas with nanomolar efficiency but also impact the human immune machinery. This conclusion is based on the pathway hits related to immune regulation and various infectious diseases, with the target genes being localized at the RNA polymerase II transcription factor complex and in caveolae.

2.7.3. Tissue-specific localization of hub genes

The dot plot illustrated in Fig. 14 delineates the outcomes of single-cell expression analysis conducted on the pivotal hub genes. This sub-study primarily aimed to scrutinize the tissue localization of these hub

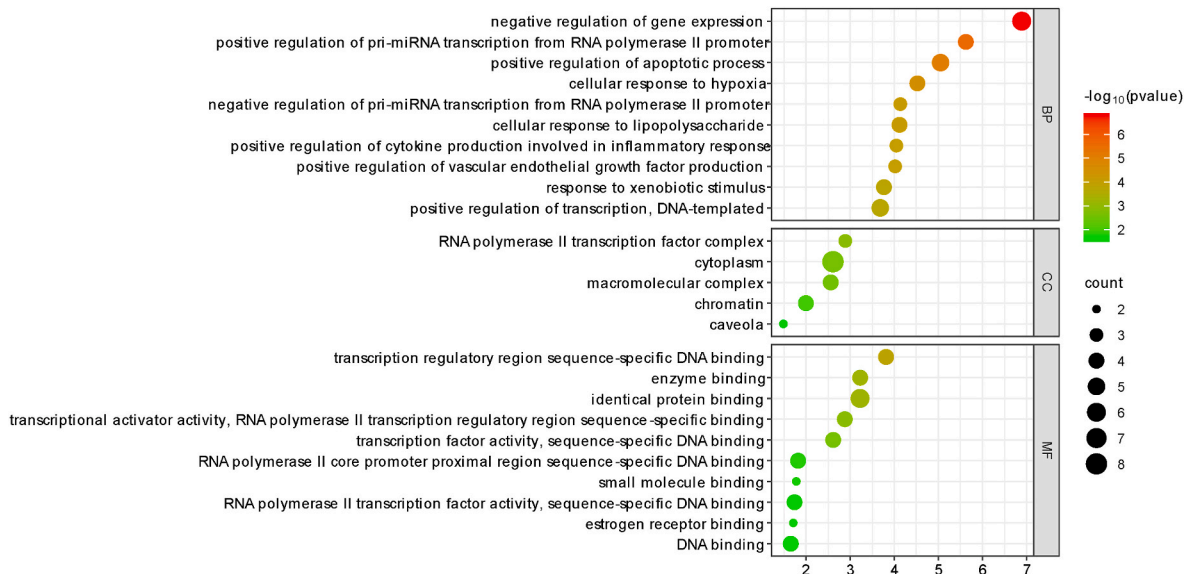


Fig. 12. Gene ontologies bubble enrichment plots depicting BP, CC and MF terms.

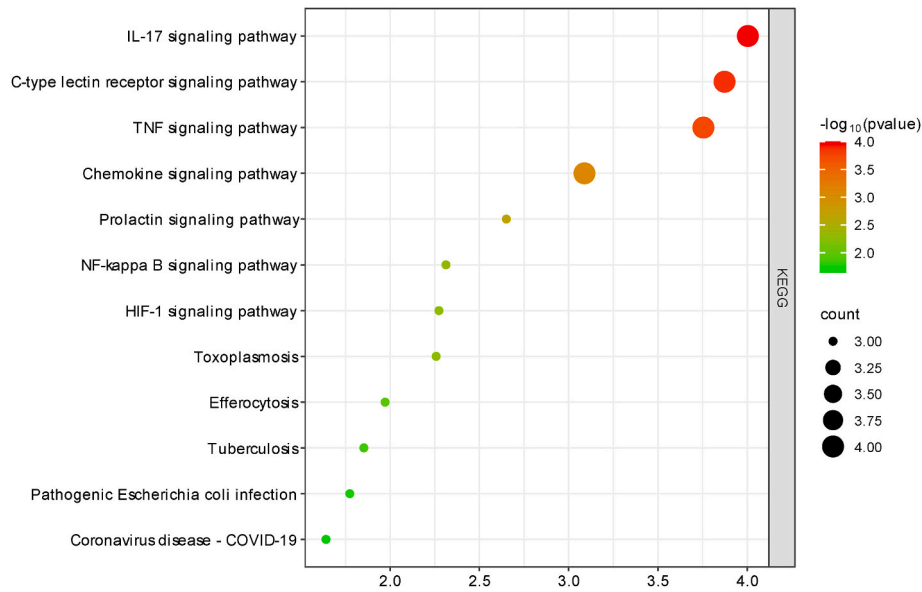


Fig. 13. Kegg pathway enrichment bubble plot.

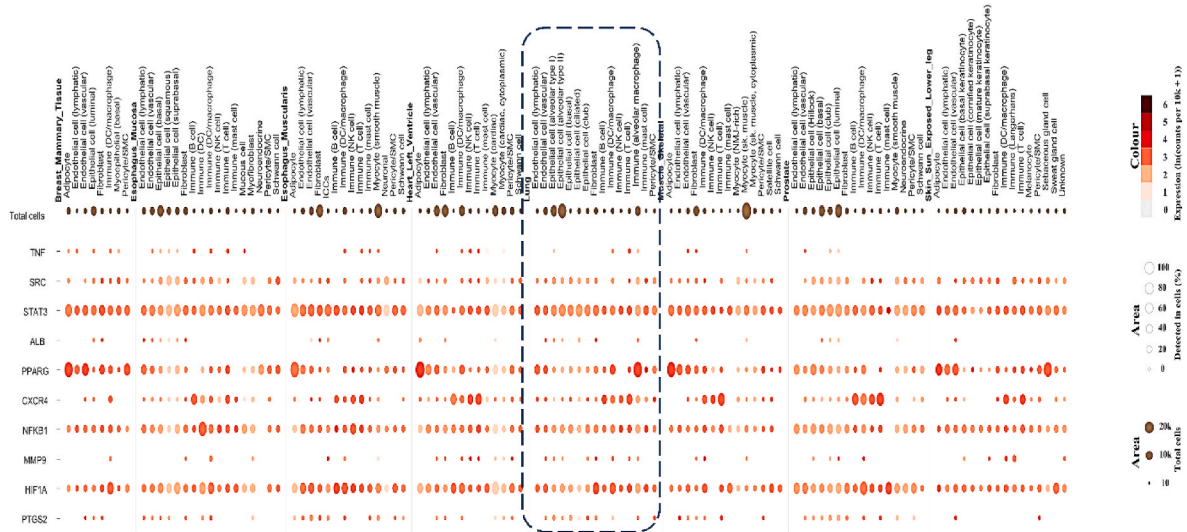


Fig. 14. Tissue localization of Hub genes.

genes, particularly focusing on specific target tissues such as the lung and esophagus muscularis, owing to their significant roles during TB infections and invasion. Notably, the genes exhibited robust expression levels within these target tissues, indicating their potential as viable targets for therapeutic interventions aimed at ameliorating the disease.

2.7.4. Tripartite network construction and analysis

The tripartite network illustrated in Fig. 15a was constructed based on synthesized compound target prediction information, hub genes, and their corresponding KEGG pathways. This network aids in the visualization of the interactions of the compounds with their corresponding hub genes to modulate a disease pathway. In this network, **NFKB1**, **TNF**, and **SRC** were found to have the greatest degree, thus ensuring that these targets among the hub genes are crucial for bridging the synthetic compounds to their ultimate goal of modulating a biological pathway. For these studies, we were more interested in exploring the probable mechanism of action of the synthesized compounds in the context of TB disease, such that edges they were connected to the tuberculosis nodes were extracted to understand influential genes and the number of

instances where the compounds were making connections with them. The hub genes **NFKB1**, **TNF**, and **SRC** with the greatest degree of interaction were found to interact with the tuberculosis pathway (illustrated in Fig. 15b).

The persistence and high communicability of *Mtb* infections underscore resistance to host immune responses. NF- κ B is a member of the family of inducible transcription factors that play a role in the mediation of adaptive and innate immunity [45]. Normally, NF- κ B is in the latent phase when it is bound to the IKB protein inside the cytoplasm of a resting cell. Once the NF- κ B is free, it accumulates in the nucleus, where it initiates the transcription of its target genes. These genes include TNF- α , chemokines, and macrophage activation molecules [46] *Mtb* infection requires proinflammatory mediators to form granulomas via tumor necrosis factor- α (TNF- α), interleukin-12 and interleukin-10. NF- κ B is associated with the release of various cytokines, such as IL-1, IL-6, IL-12, and TNF- α , via the promotion of M1 and M2 macrophage activation [47].

NF- κ B1 mediates the containment of bacterial loads within granulomas by regulating the transcription of genes involved in inflammation,

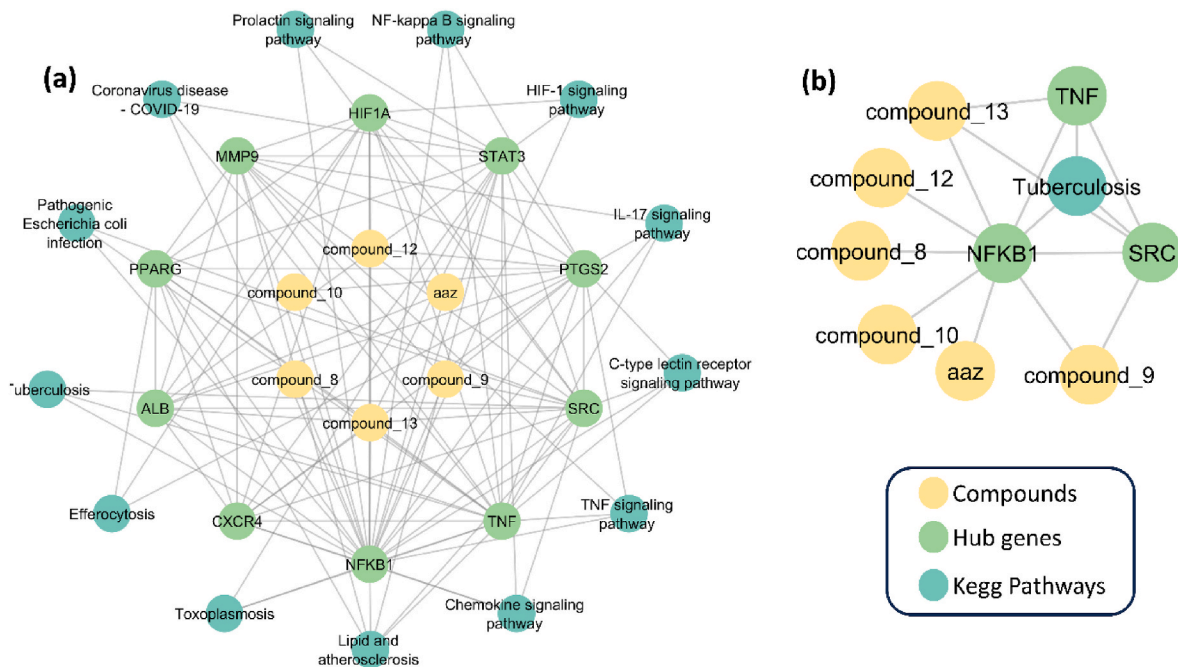


Fig. 15. Graph-based modeling for identifying MOA of synthetic compounds, a) Tripartite network interconnecting synthesized compounds with hub genes and associated KEGG pathways, and b) hierarchical network demonstrating the involvement of the synthesized comp.

such as TNF- α and chemokines. This pathway is crucial for restricting Mtb growth, and balanced NF- κ B activation is essential for preventing excessive inflammation. The negative feedback mechanism involving I κ B α and A20 attenuates NF- κ B signaling, maintaining this balance. Optimally adjusting this balance can enhance T-cell-mediated responses, potentially leading to the clearance of the bacterial infection [48,49].

Additionally, another key target identified through systems biology analysis indicated that SRC kinase is involved in modulating TB infection. Src kinases are interlinked with NF- κ B activation by mediating

tyrosine phosphorylation of the hemITAM domain on Dectin-1, creating a docking site for Syk. This interaction initiates a series of intracellular signaling cascades that culminate in activating the transcription factor NF- κ B [48] (Fig. 16). Thus, through KEGG pathway analysis, it was concluded that the synthesized molecules were involved in activating NF- κ B through direct activation or via the SRC-mediated tyrosine phosphorylation pathway.

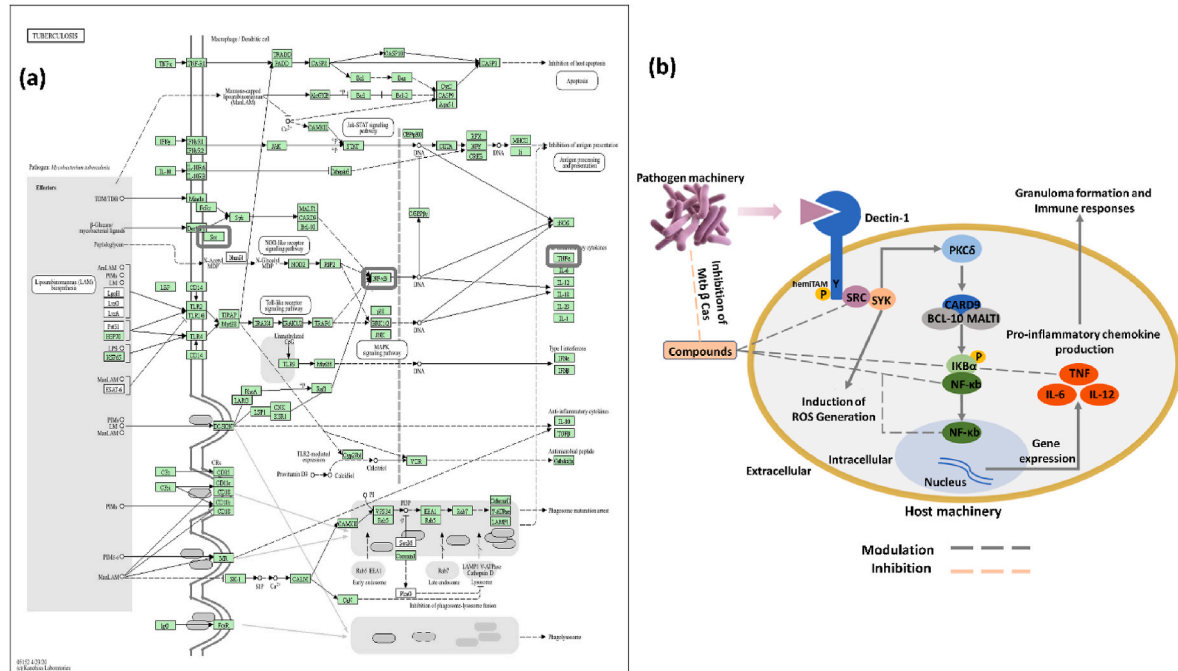


Fig. 16. Pathway analysis of the tuberculosis pathway, a) overall KEGG pathway of tuberculosis hit, and b) detailed pathway illustration of the involvement of SRC, NFKB1, and TNF-an in the dectin-1-mediated immune response and granuloma formation.

2.7.5. Grouping of β -CA inhibitors target genes and their associated pathways, phenotypes, and potential side effects

Jaccard similarity coefficients were calculated for synthesized compounds 9, 8, 10, 12, and 13 and AAZ for their corresponding target genes, pathways, and phenotypes with the CTD database using disease ID as a mapping index. The results for Jaccard similarity coefficients are illustrated as heatmaps, and the occurrence of the terms is shown by a word cloud where the size is proportional to its occurrence. The synthetic compounds had moderate similarity for their targeted genes (Fig. 17a), whereas compounds 10 and 12 exhibited the greatest similarity among all the pairs. Another high similarity was observed between compound 10 and the standard drug AAZ. Compounds 10–13 and compound 13-AAZ exhibited decreased similarity, with a coefficient of 0.255. Moderate similarity was also observed between compounds 8–9 and 8–12. Different isoforms of carbonic anhydrases were found to be targeted by synthetic compounds, as depicted in Word Cloud. (Fig. 17e).

The Jaccard similarity plot for physical molecular descriptors, such as logP, MW, number of hydrogen bond acceptors, and number of donors (Fig. 17b), indicated that all the compounds shared high similarity. This consistency confirms that these compounds are part of a congeneric series. The dataset shows that all pairs of compounds exhibit perfect similarity regarding the pathways (Fig. 17c) and phenotypes (Fig. 17c) they are associated with. This finding implies that these compounds share identical or highly similar target pathways and phenotypes. This similarity can be attributed to the similar chemical scaffolds of the compounds and the high consistency of their physical molecular descriptors.

Additionally, the most frequently associated pathways (Fig. 17g) were associated with adaptive immunity, phagosomes, innate immunity, tuberculosis, and neutrophil degranulation, again indicating the potential role of these compounds in modulating the host's immune

response in the event of TB invasion. Many hits were identified for common phenotypes, visualized as a word cloud in Fig. 17f. The associations of pathways and phenotypes with the targeted common genes were also visualized via graph-based methods, as shown in Supplementary 1, Figs. S10 and S11, respectively.

Our analysis using the NetInfer tool revealed the top 20 predicted side effects associated with synthesized compounds (9, 8, 10, 12, 13, and AAZ). These predictions were processed through hierarchical clustering to identify clusters of similar side effects, providing a comprehensive view of potential adverse reactions (Fig. 17d and h). Among the prominent side effects identified, neurological and behavioral effects such as AMNESIA, CONVULSION, and HALLUCINATION featured prominently, indicating potential impacts on cognitive and nervous system functions. Systemic effects like DRUG TOXICITY, CARDIAC ARRHYTHMIA, and FATIGUE were also prevalent, suggesting broader physiological implications. Additionally, classifications spanned various categories, including gastrointestinal effects, cough, reproductive/endocrine impacts (mammary neoplasms), and systemic disorders (ABDOMEN PAIN, CELLULAR NECROSIS). Some side effects remained unclassified, such as death (DEAD), highlighting areas where further investigation is needed. These findings, detailed in Supplementary 1, Fig. S12, underscore the diverse range of potential side effects associated with these compounds. Understanding these predictions is crucial for assessing their safety profiles and guiding future research and development efforts toward mitigating adverse health outcomes.

3. Discussions

Herein, we have described a mechanistic interpretation model to rationally design a synthesis workflow combining novel synthetic chemistry to expand the chemical space coverage of Mtb β -CAs in search

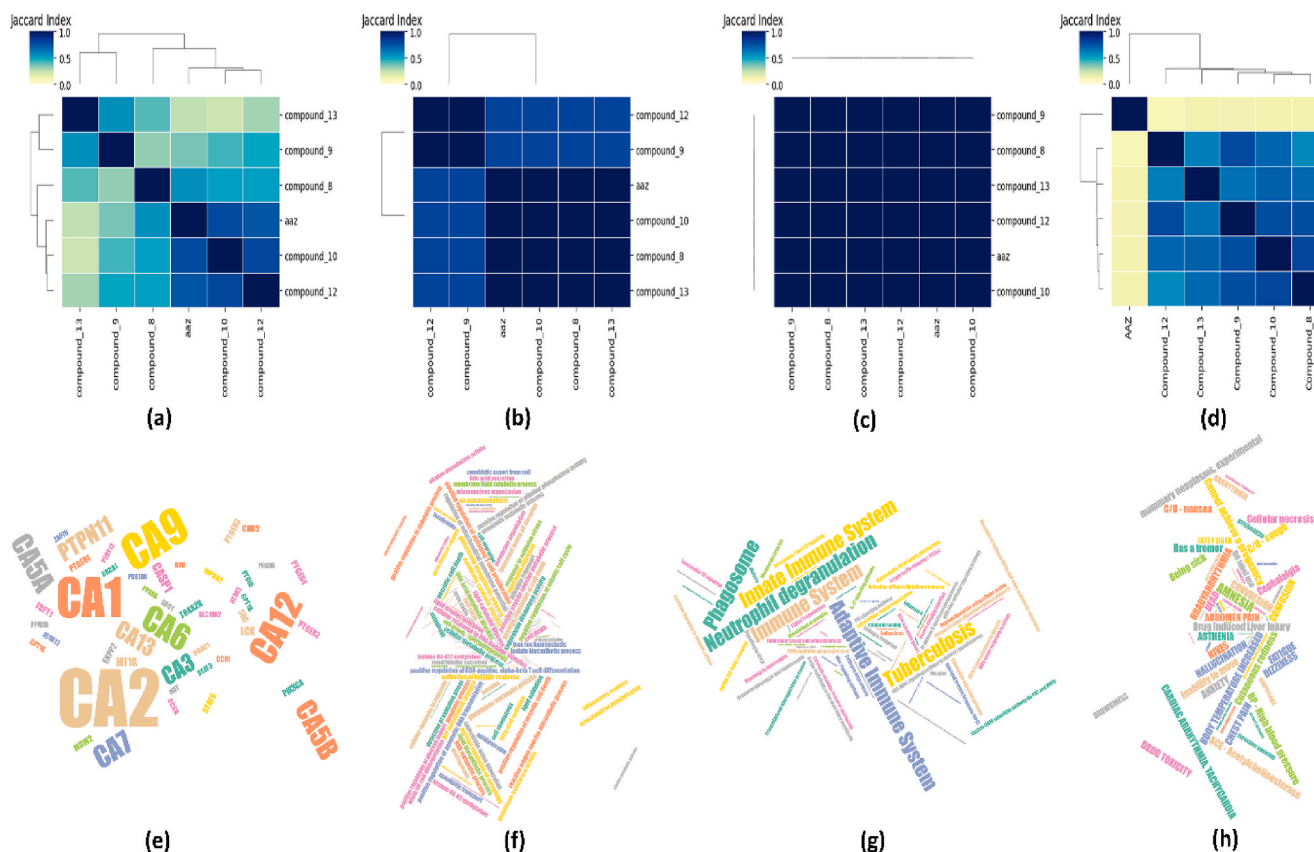


Fig. 17. Grouping of β -CA inhibitors based on similarities in their a) Interacting Genes, b) molecular descriptor (MW, logP, Number of hydrogen bond acceptors and donors), c) Pathways and phenotypes, d) side effects and word clouds by frequency of occurrence for e) Genes, f) phenotypes, g) Pathways, h) side effects.

of efficient inhibitors. The mechanistic interpretation model utilizing the random forest method performed exceptionally well, consistently maintaining high values across all the metrics. The high Cohen's kappa values further reinforce the model's reliability and robustness, with values of 0.96 for the training set and 0.84 for the test set. After validating the robustness of the mechanistic interpretation model, important chemical descriptors were identified, interpreted, and subsequently used to design new molecules. The strategically designed molecules were prioritized based on their estimated pKi values using a validated, robust ML-QSAR model. For the three target spaces of Mtb β -Ca1-3, the gradient boosting method, a tree ensemble technique, outperformed all other methods across each target space. This superior performance underscores the efficacy of gradient boosting in accurately predicting molecular activity and guiding the prioritization process. Afterward, the molecules were synthesized and evaluated for their biological activity against Mtb CAs and to understand their selectivity, such that the molecules exhibited excellent selectivity for Mtb β -CAs over hCAs and had nanomolar efficiency in inhibiting Mtb β -CAs.

We acknowledge that this study did not fully address certain methodological aspects, such as the determination of protonation states and tautomers. While our primary focus was on model development and validation, we recognize the significance of accurately representing these molecular properties in enhancing the predictive accuracy and robustness of the model. In future studies, we plan to incorporate the prediction of protonation states and tautomers using established computational tools, such as ChemAxon's Marvin or Schrödinger's Epik, to ensure a more comprehensive approach.

The synthesized compounds were evaluated for their ADME (Absorption, Distribution, Metabolism, and Excretion) characteristics and toxicity profiles to ensure efficacy and safety. Most compounds showed high GI absorption, except for molecules 12, 13, and AAZ. None exhibited blood-brain barrier (BBB) permeation or were P-glycoprotein (P-gp) substrates, which reduces the risk of drug-drug interactions and enhances cellular retention. Metabolism profiling indicated selective inhibition of CYP enzymes, with compounds 1 and 2 inhibiting CYP1A2, and compounds 3, 4, 5, 6, and 9 inhibiting CYP2C19. Toxicological analysis showed minimal concerns for most compounds, though compounds 8, 13, and AAZ presented specific risks requiring cautious handling. The ADMET properties and toxicity profiles of the compounds offer crucial insights into their potential as therapeutic agents. The high gastrointestinal (GI) absorption and limited blood-brain barrier (BBB) permeability suggest that these compounds are likely effective for treating peripheral conditions while minimizing the risk of central nervous system side effects. Additionally, the selective inhibition of CYP enzymes by these compounds points to potential drug interactions, which underscores the importance of careful management when used in conjunction with other medications. It is essential to address specific toxicity risks associated with certain compounds in subsequent development phases to ensure their safety and efficacy. Notably, the presence of efflux pumps like P-gp, which can lead to drug resistance in bacteria and host cells, highlights the need for strategies to circumvent this pathway to maintain drug efficacy. Moreover, compounds such as 8 and 13, which have shown toxicity concerns, warrant further investigation through in vivo studies to better understand their safety profiles. This comprehensive approach will contribute to the development of safer and more effective therapeutic candidates.

Molecular docking was performed to gain detailed insights into the binding mechanisms at the molecular level and to correlate the docking scores with experimental pKi values for Mtb β -CA 1, 2, and 3. Correlation coefficients were moderate for Mtb β -CA 1 ($r = 0.52$) and Mtb β -CA 2 ($r = 0.47$) but very weak for Mtb β -CA 3 ($r = 0.03$), likely due to structural inaccuracies inherent with alpha fold structures. The binding mode analysis of compounds 2, 4, and 5 showed their interaction with the catalytic triad of Mtb β -CAs, effectively inhibiting the enzyme's activity. The extended ring systems in compounds 4 and 5 increased hydrophobic interactions, stabilizing their binding poses while the zinc-binding group

anchored within the catalytic triad of Mtb β -CAs.

Our KEGG pathway analysis (p-value <0.05) highlighted the involvement of target genes of synthesized compounds in key signaling pathways, including IL-17, C-type lectin receptor, TNF, chemokine, prolactin, and NF- κ B. Pathways related to toxoplasmosis, efferocytosis, tuberculosis, pathogenic E. coli infection, and COVID-19 diseases. This underscores the dual role of the synthesized molecules in targeting Mtb and modulating the host immune response. Single-cell expression analysis showed strong expression of these hub genes in lung and esophagus tissues, which are crucial for TB infection, emphasizing their potential as therapeutic targets for TB.

The tripartite network analysis revealed critical interactions between synthesized compounds, hub genes, and KEGG pathways. **NFKB1**, **TNF**, and **SRC** emerged as key nodes central to linking synthetic compounds to disease pathways, particularly tuberculosis. NF- κ B, a key transcription factor in adaptive and innate immunity, activates genes crucial for granuloma formation and containment of Mtb. It remains latent in the cytoplasm, bound to I κ B proteins, and translocates to the nucleus upon activation, initiating TNF- α , IL-1, IL-6, IL-12, and chemokines transcription. Balanced NF- κ B activation, regulated by I κ B α and A20, is vital for controlling bacterial loads and preventing excessive inflammation. Additionally, SRC kinase modulates TB infection by mediating NF- κ B activation through tyrosine phosphorylation pathways.

Several natural compounds have been reported to exhibit immunomodulatory properties and the capacity to down-regulate the NF- κ B signaling pathway. Flavonoids, such as anthocyanins, can also control neuroinflammation and modulate the NF- κ B pathway [50,51]. Terpenoids, like betulinic acid and oleanolic acid, and phytosterols have demonstrated immunomodulatory effects and the potential to influence signaling pathways related to inflammation, immunity, and cell growth [50]. Natural compounds from bacterial sources, including lipopeptides like amphotericins, polymyxins, teicoplanins, and bacitracin, are well-known for modulating cytokine production by targeting the NF- κ B axis [52]. These findings highlight the therapeutic potential of natural products as immunomodulators and suggest their utility in mitigating adverse effects associated with the dysregulation of the NF- κ B pathway.

Further, we performed Hierarchical clustering of synthesized compounds based on target genes, pathways, phenotypes, and side effects to understand their similarity. Jaccard similarity coefficients indicated that compounds 10 and 12, and compound 10 with standard drug AAZ, exhibited the highest similarity. Physical molecular descriptors like logP, molecular weight, and hydrogen bond acceptors and donors confirmed these compounds belong to a congeneric series. Pathway and phenotype analyses showed perfect similarity, with frequently associated pathways including adaptive immunity, phagosomes, innate immunity, tuberculosis, and neutrophil degranulation, highlighting their role in modulating the host's immune response against TB. Hierarchical clustering for potential side effects identified groups of similar adverse reactions. Notable side effects included neurological and behavioral impacts such as **amnesia**, **convulsion**, and **hallucination**, as well as systemic effects like drug toxicity, **cardiac arrhythmia**, and **fatigue**. Other side effects spanned gastrointestinal, reproductive/endocrine, and systemic disorders. Some side effects, like death, remained unclassified, indicating areas for further investigation. These findings emphasize the importance of understanding the potential side effects of these compounds.

Our analysis highlights that synthesized molecules targeting Mtb β -CAs demonstrate potent anti-tuberculosis activity and modulate key immune pathways, positioning them as promising therapeutic candidates. Hub genes like **NFKB1**, **TNF**, and **SRC** play critical roles in these pathways, suggesting they are pivotal targets for enhancing immune responses and improving TB treatment outcomes. This study provides a comprehensive framework for understanding how these compounds interact with gene targets and biological pathways in the context of TB infection.

Furthermore, our systems biology analysis reveals that these

compounds not only directly inhibit Mtb β -CAs but also enhance immunity-related genomic processes. This dual action represents a departure from traditional TB therapies like isoniazid and rifampicin, which target specific bacterial enzymes. By bolstering both adaptive and innate immune responses, these compounds may effectively counteract Mtb's immune evasion mechanisms, offering potential benefits for immunocompromised patients and addressing drug resistance issues in TB treatment.

4. Conclusion

In conclusion, we have developed several machine learning models, including the mechanistic interpretation models and ML-QSAR models, to guide and prioritize synthesizing rationally designed molecules. The Biological evaluation revealed compounds with better selectivity for Mtb β -CAs over human CAs in the nanomolar range across the three isoforms of Mtb β -CA1-3. To further understand the probable mechanism of action of compounds other than those targeting the pathogen proteome, a systems biology approach was taken, and the results indicated immune system modulation during tuberculous infection. Thus, in this comprehensive study, we explained the mechanism of action of the tested molecules with a combination of cheminformatics, synthetic chemistry, biological evaluation, and systems biology approaches. Our findings offer promising insights for the development of next-generation TB therapies, as paving the way for the design of molecules that target not only the pathogen proteome but also enhance the host immune response, thus providing a comprehensive and effective approach to combatting TB.

Experimental part

Chemistry

Anhydrous solvents and all reagents were purchased from Merck, Fluorochem, and TCI. All reactions involving air- or moisture-sensitive compounds were performed under a nitrogen atmosphere using dried glassware and syringe techniques to transfer the solutions. Nuclear magnetic resonance (^1H NMR, ^{13}C NMR, ^{19}F NMR) spectra were recorded using a Bruker Advance III 400 MHz spectrometer in D_2O - d_2 , CDCl_3 - d_1 , and $\text{DMSO}-d_6$. Chemical shifts are reported in parts per million (ppm), and the coupling constants (J) are expressed in Hertz (Hz). The splitting patterns are designated as follows: s, singlet; d, doublet; t, triplet; q, quadruplet; m, multiplet; dd, doublet of doublets. The assignment of exchangeable protons was confirmed by the addition of D_2O . Analytical thin-layer chromatography (TLC) was carried out on silica gel F-254 plates (Sigma–Aldrich). Flash chromatography purifications were performed on Sigma–Aldrich silica gel 60 (230–400 mesh ASTM) as the stationary phase, and EtOAc/n-hexane or MeOH/DCM were used as eluents. Melting points (mp) were measured in open capillary tubes with a Gallenkamp MPD350.BM3.5 apparatus and were uncorrected. The solvents used for mass spectrometry analysis were acetone and acetonitrile (Chromasolv grade), which were purchased from Sigma–Aldrich (Milan, Italy), and mQ water (18 M Ω cm), which was obtained from Millipore's Simplicity system (Milan, Italy). MS analysis was carried out using a Varian 500-MS ion trap system (Palo Alto, CA, USA) equipped with two Prostar 210 pumps, a Prostar 410 autosampler, and an electrospray source (ESI) operating in negative ion mode. Stock solutions of analytes were prepared in acetone at 1.0 mg mL^{-1} and stored at 4 °C. Working solutions of each analyte were freshly prepared by diluting stock solutions in a mixture of mQ water/acetonitrile 1:1 (v/v) up to a concentration of 1.0 $\mu\text{g mL}^{-1}$. The mass spectra of each analyte were acquired by introducing the working solution via a syringe pump at 10 $\mu\text{L min}^{-1}$. The raw data were collected and processed by Varian Workstation Vers. 6.8 software. According to elemental analysis results, all compounds reported here are of >96 % purity.

General synthesis of products 1-13

Procedure A. NaOH (5 M, 0.98 equiv.) and CS_2 (1.2 equiv.) were added to a solution of the proper intermediate (0.15 g, 1 equiv.) in MeOH (5 mL) at 0 °C, and then the solution was stirred at rt for 1 h. The solution was evaporated under a vacuum, and the resulting crude product was suspended in acetone and filtered, yielding a white powder.

Procedure B. NaOH (5 M, 1.98 equiv.) and CS_2 (1.2 equiv.) were added to a solution of the proper intermediate (0.15 g, 1 equiv.) in MeOH (5 mL) at 0 °C, and then the solution was stirred at rt for 1 h. The solution was evaporated under a vacuum, and the resulting crude product was suspended in acetone and filtered, yielding a white powder.

Sodium (4-sulfamoylbenzyl)carbamdithioate (1)

Compound **1** was obtained according to the general procedure A earlier reported with 4-sulfamoylbenzyl amine (**14**) as starting material. Yield 97 %; silica gel R_f 0.24 (TFA/MeOH/DCM 7/9/84 % v/v); δ_H (400 MHz, $\text{DMSO}-d_6$): 8.61 (t, $J = 6.1$ Hz, 1H, NHCS_2^-), 7.75 (d, $J = 8.2$ Hz, 2H, 2 x Ar-H), 7.43 (d, $J = 8.2$ Hz, 2H, 2 x Ar-H), 7.28 (s, 2H, exchange with D_2O , SO_2NH_2), 4.79 (d, $J = 6.1$ Hz, 2H, CH_2); δ_C (100 MHz, D_2O): 213.7, 144.7, 139.9, 128.3, 126.8, 51.2; m/z (ESI negative): 261.1 [$\text{M} - \text{H}$] $^-$; Elemental analysis: calculated C, 33.79; H, 3.19; N, 9.85; found C, 33.78; H, 3.17; N, 9.86.

Sodium (4-sulfamoylphenethyl)carbamdithioate (2)

Compound **2** was obtained according to the general procedure A earlier reported with 4-sulfamoylphenethyl amine (**15**) as starting material. Yield 87 %; silica gel R_f 0.20 (TFA/MeOH/DCM 7/9/84 % v/v); δ_H (400 MHz, $\text{DMSO}-d_6$): 8.61 (t, $J = 6.1$ Hz, 1H, NHCS_2^-), 7.75 (d, $J = 8.2$ Hz, 2H, 2 x Ar-H), 7.43 (d, $J = 8.2$ Hz, 2H, 2 x Ar-H), 7.28 (s, 2H, exchange with D_2O , SO_2NH_2), 4.79 (d, $J = 6.1$ Hz, 2H, CH_2); δ_C (100 MHz, D_2O): 210.6, 146.0, 139.7, 130.6, 126.7, 48.9, 34.5; m/z (ESI negative): 275.0 [$\text{M} - \text{H}$] $^-$; Elemental analysis: calculated C, 36.23; H, 3.72; N, 9.39; found C, 36.22; H, 3.73; N, 9.37.

Sodium 4-(4-sulfamoylphenyl)piperazine-1-carbodithioate (3)

Compound **3** was obtained according to the general procedure A earlier reported with 4-(piperazin-1-yl)benzenesulfonamide (**18**) as starting materials. Yield 91 %; silica gel R_f 0.21 (TFA/MeOH/DCM 7/9/84 % v/v); δ_H (400 MHz, $\text{DMSO}-d_6$): 7.66 (d, $J = 8.6$ Hz, 2H, 2 x Ar-H), 7.06 (s, 2H, exchange with D_2O , SO_2NH_2), 7.04 (d, $J = 8.6$ Hz, 2H, 2 x Ar-H), 4.50 (m, 4H, 2 x CH_2), 3.30 (m, 4H, 2 x CH_2); δ_C (100 MHz, D_2O): 209.4, 153.9, 130.7, 128.2, 115.4, 50.8, 47.4; m/z (ESI negative): 316.1 [$\text{M} - \text{H}$] $^-$; Elemental analysis: calculated C, 38.93; H, 4.16; N, 12.38; found C, 38.94; H, 4.14; N, 12.33.

Sodium 4-(2-fluoro-4-sulfamoylphenyl)piperazine-1-carbodithioate (4)

Compound **4** was obtained according to the general procedure A earlier reported with 3-fluoro-4-(piperazin-1-yl)benzenesulfonamide (**19**) as starting material. Yield 93 %; silica gel R_f 0.21 (TFA/MeOH/DCM 7/9/84 % v/v); δ_H (400 MHz, $\text{DMSO}-d_6$): 7.56 (m, 2H, 2 x Ar-H), 7.27 (s, 2H, exchange with D_2O , SO_2NH_2), 7.21 (m, 1H, Ar-H), 5.54 (m, 4H, 2 x CH_2), 3.09 (m, 4H, 2 x CH_2); δ_C (100 MHz, D_2O): 209.9, 153.7 ($J_{\text{C-F}}^1 = 251.4$ Hz), 143.4 ($J_{\text{C-F}}^3 = 7.7$ Hz), 136.3 ($J_{\text{C-F}}^3 = 7.7$ Hz), 123.6 ($J_{\text{C-F}}^2 = 2.8$ Hz), 120.6, 114.9 ($J_{\text{C-F}}^2 = 26.7$ Hz), 51.0, 50.1; δ_F (400 MHz, D_2O): 119.5; m/z (ESI negative): 334.1 [$\text{M} - \text{H}$] $^-$; Elemental analysis: calculated C, 36.97; H, 3.67; N, 11.76; found C, 36.98; H, 3.66; N, 11.75.

Sodium 4-(4-sulfamoylphenyl)-1,4-diazepane-1-carbodithioate (5)

Compound **5** was obtained according to the general procedure A earlier reported with 4-(1,4-diazepan-1-yl)benzenesulfonamide (**20**) as starting material. Yield 44 %; silica gel R_f 0.20 (TFA/MeOH/DCM 7/9/84 % v/v); δ_H (400 MHz, $\text{DMSO}-d_6$): 7.57 (d, $J = 8.1$ Hz, 2H, 2 x Ar-H), 6.94 (s, 2H, exchange with D_2O , SO_2NH_2), 6.84 (d, $J = 8.1$ Hz, 2H, 2 x Ar-H), 4.40 (t, $J = 4.7$ Hz, 2H, CH_2), 3.95 (t, $J = 5.5$ Hz, 2H, CH_2), 3.64 (t, $J = 4.7$ Hz, 2H, CH_2), 3.48 (t, $J = 5.5$ Hz, 2H, CH_2), 2.02 (m, 2H, CH_2);

δ_C (100 MHz, D_2O): 209.9, 151.3, 128.6, 127.1, 112.2, 54.6, 53.5, 49.1, 48.2, 22.6; m/z (ESI negative): 330.0 $[M - H]^-$; Elemental analysis: calculated C, 40.78; H, 4.56; N, 11.89; found C, 40.80; H, 4.54; N, 11.88.

Sodium 4-(2-fluoro-4-sulfamoylphenyl)-1,4-diazepane-1-carbodithioate (6)

Compound **6** was obtained according to the general procedure A earlier reported with 4-(1,4-diazepan-1-yl)-3-fluorobenzenesulfonamide (**21**) as starting material. Yield 15 %; silica gel R_f 0.12 (TFA/MeOH/DCM 7/9/84 % v/v); δ_H (400 MHz, DMSO- d_6): 7.45 (m, 2H, 2 x Ar-H), 7.10 (m, 3H, exchange with D_2O , SO_2NH_2 + Ar-H), 4.48 (m, 2H, CH_2), 4.14 (m, 2H, CH_2), 3.62 (m, 2H, CH_2), 3.47 (m, 2H, CH_2), 1.95 (m, 2H, CH_2); δ_C (100 MHz, D_2O): 209.8, 153.3, 141.9 ($J^3_{C-F} = 7.3$ Hz), 131.0 ($J^3_{C-F} = 8$ Hz), 123.8, 118.0 ($J^4_{C-F} = 4.3$ Hz), 115.5 ($J^2_{C-F} = 26.4$ Hz), 55.1, 53.5, 51.6, 51.5, 51.4, 25.1; δ_f (400 MHz, D_2O): 122.1; m/z (ESI negative): 348.0 $[M - H]^-$; Elemental analysis: calculated C, 38.80; H, 4.07; N, 11.31; found C, 38.82; H, 4.09; N, 11.29.

Sodium 4-(4-sulfamoylbenzoyl)piperazine-1-carbodithioate (7)

Compound **7** was obtained according to the general procedure A earlier reported with 4-(4-Sulfamoylbenzoyl)piperazin-1-ium chloride (**24**) as starting material. Yield 69 %; silica gel R_f 0.23 (TFA/MeOH/DCM 7/9/84 % v/v); δ_H (400 MHz, DMSO- d_6): 7.91 (d, $J = 7.8$ Hz, 2H, 2 x Ar-H), 7.65 (d, $J = 7.8$ Hz, 2H, 2 x Ar-H), 7.44 (s, 2H, exchange with D_2O , SO_2NH_2), 4.49 (s, 2H, CH_2), 4.38 (s, 2H, CH_2), 3.61 (m, 2H, CH_2), 3.29 (m, 2H, CH_2); δ_C (100 MHz, D_2O): 205.3, 171.6, 143.7, 138.9, 128.6, 127.2, 47.5, 42.7; m/z (ESI negative): 344.1 $[M - H]^-$; Elemental analysis: calculated C, 39.23; H, 3.84; N, 11.44; found C, 39.22; H, 3.85; N, 11.45.

Sodium 4-(2-oxo-2-((4-sulfamoylphenyl)amino)ethyl)piperazine-1-carbodithioate (8)

Compound **8** was obtained according to the general procedure B earlier reported with 1-(2-Oxo-2-((4-sulfamoylphenyl)amino)ethyl)piperazine-1,4-diium chloride (**28**) as starting material. Yield 69 %; silica gel R_f 0.13 (TFA/MeOH/DCM 7/9/84 % v/v); δ_H (400 MHz, DMSO- d_6): 10.20 (s, 1H, exchange with D_2O , CONH), 7.87 (d, $J = 8.4$ Hz, 2H, 2 x Ar-H), 7.78 (d, $J = 8.4$ Hz, 2H, 2 x Ar-H), 7.30 (s, 2H, exchange with D_2O , SO_2NH_2), 4.29 (t, $J = 4.5$ Hz, 4H, 2 x CH_2), 3.21 (s, 2H, CH_2), 2.48 (t, $J = 4.5$ Hz, 4H, 2 x CH_2); δ_C (100 MHz, D_2O): 209.3, 171.6, 141.6, 137.7, 127.7, 121.9, 60.9, 52.8, 51.2; m/z (ESI negative): 373.4 $[M - H]^-$; Elemental analysis: calculated C, 39.38; H, 4.32; N, 14.13; found C, 39.40; H, 4.33; N, 14.14.

Sodium 4-(4-sulfamoylbenzyl)piperazine-1-carbodithioate (9)

Compound **9** was obtained according to the general procedure B earlier reported with 1-(4-Sulfamoylbenzyl)piperazine-1,4-diium chloride (**31**) as starting material. Yield 0.83 %; silica gel R_f 0.18 (TFA/MeOH/DCM 7/9/84 % v/v); δ_H (400 MHz, DMSO- d_6): 7.81 (d, $J = 7.7$ Hz, 2H, 2 x Ar-H), 7.53 (d, $J = 7.7$ Hz, 2H, 2 x Ar-H), 7.25 (s, 2H, exchange with D_2O , SO_2NH_2), 4.37 (m, 4H, 2 x CH_2), 3.58 (s, 2H, CH_2), 2.33 (m, 4H, 2 x CH_2); δ_C (100 MHz, D_2O): 209.8, 142.7, 141.1, 131.5, 126.3, 61.5, 52.4, 50.8; m/z (ESI negative): 330.1 $[M - H]^-$; Elemental analysis: calculated C, 40.78; H, 4.56; N, 11.89; found C, 40.76; H, 4.58; N, 11.88.

Sodium 4-((4-sulfamoylbenzyl)carbamoyl)piperazine-1-carbodithioate (10)

Compound **10** was obtained according to the general procedure A earlier reported with 4-((4-Sulfamoylbenzyl)carbamoyl)piperazin-1-ium chloride (**36**) as starting material. Yield 82 %; silica gel R_f 0.20 (TFA/MeOH/DCM 7/9/84 % v/v); δ_H (400 MHz, DMSO- d_6): 7.78 (d, $J = 7.9$ Hz, 2H, 2 x Ar-H), 7.46 (d, $J = 7.9$ Hz, 2H, 2 x Ar-H), 7.31 (s, 2H, exchange with D_2O , SO_2NH_2), 7.22 (t, $J = 5.8$ Hz, 1H, exchange with D_2O , NHCON), 4.34 (m, 6H, 3 x CH_2), 3.32 (t, $J = 5.2$ Hz, 4H, 2 x CH_2); δ_C (100 MHz, D_2O): 210.1, 159.9, 146.0, 140.4, 128.1, 126.7, 51.0, 44.2, 43.3; m/z (ESI negative): 373.4 $[M - H]^-$; Elemental analysis: calculated C, 40.78; H, 4.56; N, 11.89; found C, 40.76; H, 4.58; N, 11.88.

Sodium 4-((4-sulfamoylphenyl)carbamoyl)piperazine-1-carbodithioate (11)

Compound **11** was obtained according to the general procedure A earlier reported with 4-((4-Sulfamoylphenyl)carbamoyl)piperazin-1-ium chloride (**37**) as starting material. Yield 97 %; silica gel R_f 0.25 (TFA/MeOH/DCM 7/9/84 % v/v); δ_H (400 MHz, DMSO- d_6): 7.78 (d, $J = 7.4$ Hz, 2H, 2 x Ar-H), 7.41 (d, $J = 7.4$ Hz, 2H, 2 x Ar-H), 7.33 (s, 2H, exchange with D_2O , SO_2NH_2), 6.67 (t, $J = 5.4$ Hz, 1H, exchange with D_2O , NHCON), 4.32 (m, 4H, 2 x CH_2), 3.27 (m, 6H, 3 x CH_2), 2.83 (t, $J = 6.7$ Hz, 2H, CH_2); δ_C (100 MHz, D_2O): 210.1, 159.8, 146.1, 139.7, 130.6, 126.5, 50.8, 43.2, 41.6, 36.0; m/z (ESI negative): 387.0 $[M - H]^-$; Elemental analysis: calculated C, 40.96; H, 4.67; N, 13.65; found C, 40.97; H, 4.65; N, 13.67.

Sodium 4-((4-sulfamoylphenyl)carbamothioyl)piperazine-1-carbodithioate (12)

Compound **12** was obtained according to the general procedure A earlier reported with 4-((4-Sulfamoylphenyl)carbamothioyl)piperazin-1-ium chloride (**40**) as starting material. Yield 79 %; silica gel R_f 0.29 (TFA/MeOH/DCM 7/9/84 % v/v); δ_H (400 MHz, DMSO- d_6): 9.57 (s, 1H, exchange with D_2O , NHCSN), 7.74 (d, $J = 8.0$ Hz, 2H, 2 x Ar-H), 7.53 (d, $J = 8.0$ Hz, 2H, 2 x Ar-H), 7.30 (s, 2H, exchange with D_2O , SO_2NH_2), 4.43 (m, 4H, 2 x CH_2), 3.93 (m, 4H, 2 x CH_2); δ_C (100 MHz, D_2O): 210.2, 180.6, 144.8, 138.7, 127.3, 126.8, 50.3, 47.9; m/z (ESI negative): 375.0 $[M - H]^-$; Elemental analysis: calculated C, 36.17; H, 3.79; N, 14.06; found C, 36.18; H, 3.77; N, 14.07.

Sodium 4-((5-sulfamoyl-1,3,4-thiadiazol-2-yl)carbamoyl)piperazine-1-carbodithioate (13)

Compound **13** was obtained according to the general procedure A earlier reported with 4-((5-Sulfamoyl-1,3,4-thiadiazol-2-yl)carbamoyl)piperazin-1-ium chloride (**45**) as starting material. Yield 71 %; silica gel R_f 0.26 (TFA/MeOH/DCM 7/9/84 % v/v); δ_H (400 MHz, DMSO- d_6): 7.50 (s, 2H, exchange with D_2O , SO_2NH_2), 4.42 (m, 4H, 2 x CH_2), 3.39 (m, 4H, 2 x CH_2); δ_C (100 MHz, D_2O): 210.6, 183.0, 176.3, 163.7, 51.6, 43.6; m/z (ESI negative): 367.1 $[M - H]^-$; Elemental analysis: calculated C, 24.61; H, 2.84; N, 21.52; found C, 24.60; H, 2.86; N, 21.50.

General synthesis for 4-(piperazin-1-yl)benzenesulfonamide (18), 3-fluoro-4-(piperazin-1-yl)benzenesulfonamide (19), 4-(1,4-diazepan-1-yl)benzenesulfonamide (20) and 4-(1,4-diazepan-1-yl)-3-fluorobenzenesulfonamide (21)

Piperazine or homopiperazine (5 equiv.) was added to a solution of **16** or **17** (0.5 g, 1 equiv.) in H_2O (10 mL), and then the suspension was heated overnight at reflux. The solution was cooled, and the resulting precipitate was filtered and washed with cold H_2O , yielding a white powder.

4-(Piperazin-1-yl)benzenesulfonamide (18)

Compound **18** was obtained according to the general procedure previously reported with 4-fluorobenzenesulfonamide (**16**) and piperazine as the starting materials. Yield 91 %; m.p. 210–211 °C; silica gel R_f 0.15 (MeOH/DCM 20 % v/v); δ_H (400 MHz, DMSO- d_6): 7.49 (d, $J = 8.7$ Hz, 2H, 2 x Ar-H), 7.06 (s, 2H, exchange with D_2O , SO_2NH_2), 6.70 (d, $J = 8.7$ Hz, 2H, 2 x Ar-H), 3.09 (t, $J = 5.9$ Hz, 4H, 2 x CH_2), 2.87 (t, $J = 5.9$ Hz, 4H, 2 x CH_2).

3-fluoro-4-(piperazin-1-yl)benzenesulfonamide (19)

Compound **19** was obtained according to the general procedure previously reported with 3,4-difluorobenzenesulfonamide (**17**) and piperazine as the starting materials. Yield %; m.p. 231–233 °C; silica gel R_f 0.15 (MeOH/DCM 20 % v/v); δ_H (300 MHz, DMSO- d_6): 7.53 (dd, $J = 8.6$ Hz 2 Hz, 1H, Ar-H), 7.48 (dd, $J = 15$ Hz 2 Hz, 1H, Ar-H), 7.31 (s, 2H, exchange with D_2O , SO_2NH_2), 7.13 (dd, $J = 8.6$ Hz 8.6 Hz 1H, Ar-H), 3.03 (m, 4H, 2 x CH_2), 2.83 (m, 4H, 2 x CH_2).

4-(1,4-Diazepan-1-yl)benzenesulfonamide (20)

Compound **20** was obtained according to the general procedure previously reported with 4-fluorobenzenesulfonamide (**16**) and homopiperazine as the starting materials. Yield 78 %; m.p. 215–217 °C; silica gel R_f 0.15 (MeOH/DCM 20 % v/v); δ_H (400 MHz, DMSO- d_6): 7.56 (d, J = 9.0 Hz, 2H, 2 x Ar-H), 6.95 (s, 2H, exchange with D₂O, SO₂NH₂), 6.77 (d, J = 9.0 Hz, 2H, 2 x Ar-H), 3.59 (t, J = 5.9 Hz, 2H, CH₂), 3.51 (t, J = 5.2 Hz, 2H, CH₂), 2.84 (t, J = 5.2 Hz, 2H, CH₂), 2.60 (t, J = 5.9 Hz, 2H, CH₂), 1.76 (m, 2H, CH₂).

4-(1,4-Diazepan-1-yl)-3-fluorobenzenesulfonamide (21)

Compound **21** was obtained according to the general procedure previously reported with 3,4-difluorobenzenesulfonamide (**17**) and homopiperazine as the starting materials. Yield 84 %; m.p. 240–243 °C; silica gel R_f 0.17 (MeOH/DCM 20 % v/v); δ_H (400 MHz, DMSO- d_6): 7.53 (m, 1H, Ar-H), 7.48 (m, 1H, Ar-H), 7.41 (s, 2H, exchange with D₂O, SO₂NH₂), 7.18 (m, 1H, Ar-H), 3.68 (t, J = 5.4 Hz, 2H, CH₂), 3.46 (t, J = 5.0 Hz, 2H, CH₂), 2.92 (t, J = 5.0 Hz, 2H, CH₂), 2.71 (t, J = 5.4 Hz, 2H, CH₂), 1.74 (m, 2H, CH₂).

Synthesis of *t*-butyl 4-(4-sulfamoylbenzoyl)piperazine-1-carboxylate (23)

PyBOP (1.2 equiv.), *t*-butyl piperazine-1-carboxylate (1.5 equiv.) and DIPEA (3 equiv.) were added to a solution of **22** (0.5 g, 1 equiv.) in anhydrous DMF (3 mL) at 0 °C under a nitrogen atmosphere, and then the resulting suspension was stirred overnight at RT. Slush and 12 M HCl were added until the pH reached 4, and the resulting precipitate was filtered and washed with cold H₂O to yield a white powder.

Yield 87 %; m.p. 188–189 °C; silica gel R_f 0.24 (MeOH/DCM 10 % v/v); δ_H (400 MHz, DMSO- d_6): 7.92 (d, J = 7.8 Hz, 2H, 2 x Ar-H), 7.64 (d, J = 7.8 Hz, 2H, 2 x Ar-H), 7.51 (s, 2H, exchange with D₂O, SO₂NH₂), 3.64 (m, 2H, CH₂), 3.38 (m, H₂O overlapped, 6H, 3 x CH₂), 1.44 (s, 9H, 3 x CH₃).

Synthesis of 2-chloro-*N*-(4-sulfamoylphenyl)acetamide (26)

K₂CO₃ (1.2 equiv.) and chloroacetyl chloride (1.2 equiv.) were added to a solution of **25** (0.5 g, 1 equiv.) in acetone (15 mL) at 0 °C, and then the reaction mixture was stirred at RT for 30 min. The solvent was evaporated under vacuum, slush and 12 M HCl were added until the pH reached 2, and the resulting suspension was filtered. The obtained white powder was washed with cold H₂O.

Yield 98 %; silica gel R_f 0.19 (MeOH/DCM 10 % v/v); δ_H (400 MHz, DMSO- d_6): 10.62 (s, 1H, exchange with D₂O, CONH), 7.77 (d, J = 8.9 Hz, 2H, 2 x Ar-H), 7.73 (d, J = 8.9 Hz, 2H, 2 x Ar-H), 7.26 (s, 2H, exchange with D₂O, SO₂NH₂), 4.28 (s, 2H, CH₂).

Synthesis of *t*-butyl 4-(2-oxo-2-((4-sulfamoylphenyl)amino)ethyl)piperazine-1-carboxylate (27)

K₂CO₃ (1.2 equiv.) and *t*-butyl piperazine-1-carboxylate (1.2 equiv.) were added to a solution of **26** (0.5 g, 1 equiv.) in anhydrous DMF (5 mL) under a nitrogen atmosphere, and then the resulting suspension was stirred overnight at RT. Slush was added, and the precipitate was filtered to give a white powder.

Yield 71 %; m.p. 195–197 °C; silica gel R_f 0.14 (MeOH/DCM 6 % v/v); δ_H (400 MHz, DMSO- d_6): 10.12 (s, 1H, exchange with D₂O, CONH), 7.84 (d, J = 8.7 Hz, 2H, 2 x Ar-H), 7.79 (d, J = 8.7 Hz, 2H, 2 x Ar-H), 7.30 (s, 2H, exchange with D₂O, SO₂NH₂), 3.41 (m, 4H, 2 x CH₂), 3.23 (s, 2H, CH₂), 2.51 (m, 4H, 2 x CH₂), 1.44 (s, 9H, 3 x CH₃).

5.2.22. Synthesis of *t*-butyl 4-(4-sulfamoylbenzyl)piperazine-1-carboxylate (30)

K₂CO₃ (1.5 equiv.) and *t*-butyl piperazine-1-carboxylate (1.5 equiv.) were added to a solution of **29** (0.3 g, 1 equiv.) in anhydrous DMF (2 mL) under a nitrogen atmosphere, and then the resulting suspension was stirred overnight at RT. Slush was added, and the precipitate was filtered to give a white powder.

Yield 72 %; m.p. 201–204 °C; silica gel R_f 0.22 (MeOH/DCM 10 % v/v);

δ_H (400 MHz, DMSO- d_6): 7.80 (d, J = 8.0 Hz, 2H, 2 x Ar-H), 7.46 (d, J = 8.0 Hz, 2H, 2 x Ar-H), 7.32 (s, 2H, exchange with D₂O, SO₂NH₂), 3.50 (s, 2H, CH₂), 3.38 (t, H₂O overlapped, J = 4.4 Hz, 4H, 2 x CH₂), 2.33 (t, J = 4.4 Hz, 4H, 2 x CH₂), 1.42 (s, 9H, 3 x CH₃).

General synthesis of phenyl (4-sulfamoylbenzyl)carbamate (32) and phenyl (4-sulfamoylphenetyl)carbamate (33)

K₂CO₃ (2 equiv.) and phenylchloro formate (1.5 equiv.) were added to a suspension of **14** or **15** (3 g, 1 equiv.) in acetone (50 mL) cooled to 0 °C, after which the reaction mixture was stirred for 2 h at RT. The solvent was removed under vacuum, and the crude product was treated with a slush and 12 M HCl until the pH reached 2. The resulting suspension was filtered and washed with H₂O and Et₂O to give a white powder.

Phenyl (4-sulfamoylbenzyl)carbamate (32)

Compound **32** was obtained according to the general procedure previously reported with 4-sulfamoylbenzyl amine (**14**) as the starting material. Yield 87 %; m.p. 256–257 °C; silica gel TLC R_f 0.48 (MeOH/DCM 10 % v/v); δ_H (400 MHz, DMSO- d_6): 8.42 (t, J = 5.5 Hz, 1H, exchange with D₂O, CONH), 7.83 (d, J = 7.1 Hz, 2H, 2 x Ar-H), 7.52 (d, J = 7.1 Hz, 2H, 2 x Ar-H), 7.40 (m, 4H, Ar-H), 7.36 (d, 2H, exchange with D₂O, SO₂NH₂), 7.23 (t, J = 7.4 Hz, 1H, Ar-H), 7.15 (d, J = 7.4 Hz, 2H, Ar-H), 4.37 (d, J = 5.5 Hz, 2H, CH₂).

Phenyl (4-sulfamoylphenetyl)carbamate (33)

Compound **33** was obtained according to the general procedure previously reported with 4-sulfamoylphenetyl amine (**15**) as the starting material. Yield 94 %; m.p. 235–238 °C; silica gel TLC R_f 0.62 (MeOH/DCM 10 % v/v); δ_H (400 MHz, DMSO- d_6): 7.88 (t, J = 5.6 Hz, 1H, exchange with D₂O, CONH), 7.79 (d, J = 7.7 Hz, 2H, 2 x Ar-H), 7.46 (d, J = 7.7 Hz, 2H, 2 x Ar-H), 7.38 (d, J = 7.7 Hz, 2H, 2 x Ar-H), 7.33 (s, 2H, exchange with D₂O, SO₂NH₂), 7.21 (t, J = 7.7 Hz, 1H, Ar-H), 7.08 (d, J = 7.7 Hz, 2H, 2 x Ar-H), 3.32 (m, 2H, CH₂), 2.84 (t, J = 6.9 Hz, 2H, CH₂).

General synthesis of *t*-butyl 4-((4-sulfamoylbenzyl)carbamoyl)piperazine-1-carboxylate (34) and *t*-butyl 4-((4-sulfamoylphenetyl)carbamoyl)piperazine-1-carboxylate (35)

DIPEA (cat.) and *t*-butyl piperazine-1-carboxylate (1.5 equiv.) were added to a solution of **32** or **33** (0.5 g, 1 equiv.) in anhydrous ACN (8 mL) under a nitrogen atmosphere, and then the resulting suspension was stirred overnight at reflux. The solution was cooled, and a slush was added. Then, the obtained precipitate was filtered to obtain a white powder.

t-Butyl 4-((4-sulfamoylbenzyl)carbamoyl)piperazine-1-carboxylate (34)

Compound **34** was obtained according to the general procedure previously reported with phenyl (4-sulfamoylbenzyl)carbamate (**32**) and piperazine as the starting materials. Yield 92 %; m.p. 247–248 °C; silica gel R_f 0.29 (MeOH/DCM 8 % v/v); δ_H (400 MHz, DMSO- d_6): 7.76 (d, J = 8.5 Hz, 2H, 2 x Ar-H), 7.43 (d, J = 8.5 Hz, 2H, 2 x Ar-H), 7.32 (s, 2H, exchange with D₂O, SO₂NH₂), 7.26 (t, J = 6.0 Hz, 1H, exchange with D₂O, NHCON), 4.30 (d, J = 6.0 Hz, 2H, CH₂), 3.31 (m, 8H, 4 x CH₂), 1.42 (s, 9H, 3 x CH₃).

t-Butyl 4-((4-sulfamoylphenetyl)carbamoyl)piperazine-1-carboxylate (35)

Compound **35** was obtained according to the general procedure previously reported with phenyl (4-sulfamoylphenetyl)carbamate (**33**) and piperazine as the starting materials. Yield 88 %; m.p. 258–261 °C; silica gel R_f 0.31 (MeOH/DCM 8 % v/v); δ_H (400 MHz, DMSO- d_6): 7.77 (d, J = 8.3 Hz, 2H, 2 x Ar-H), 7.41 (d, J = 8.3 Hz, 2H, 2 x Ar-H), 7.34 (s, 2H, exchange with D₂O, SO₂NH₂), 6.73 (t, J = 5.0 Hz, 1H, exchange with D₂O, NHCON), 3.29 (m, 10H, 5 x CH₂), 2.82 (t, J = 6.2 Hz, 2H, CH₂), 1.44 (s, 9H, 3 x CH₃).

Synthesis of *t*-butyl 4-((4-sulfamoylphenyl)carbamothioyl)piperazine-1-carboxylate (39)

DIPEA (cat.) and *t*-butyl piperazine-1-carboxylate (1.5 equiv.) were added to a solution of **38** (0.5 g, 1 equiv.) in anhydrous ACN (8 mL) under a nitrogen atmosphere, after which the resulting suspension was stirred overnight at RT. Slush was added, and the obtained precipitate was filtered to give a white powder.

Yield 94 %; m.p. 277–280 °C; silica gel R_f 0.26 (EtOAc/hexane 50 % v/v); δ_H (400 MHz, DMSO- d_6): 9.60 (s, 1H, exchange with D₂O, NHCSN), 7.73 (d, J = 8.4 Hz, 2H, 2 x Ar-H), 7.49 (d, J = 8.4 Hz, 2H, 2 x Ar-H), 7.29 (s, 2H, exchange with D₂O, SO₂NH₂), 3.92 (t, J = 5.2 Hz, 4H, 2 x CH₂), 3.45 (t, J = 5.2 Hz, 4H, 2 x CH₂), 1.44 (s, 9H, 3 x CH₃).

Synthesis of 5-amino-1,3,4-thiadiazole-2-sulfonamide (42)

HCl (12 M) was added to a suspension of **41** (1.5 g, 1 equiv.) in EtOH, and the mixture was stirred overnight at reflux. The solvent was removed under vacuum, the crude product was suspended in H₂O (10 mL), and 5 M NaOH was added until the pH reached 6. The resulting precipitate was filtered and washed with cold H₂O to afford a white powder.

Yield 86 %; m.p. 218–229 °C; silica gel TLC R_f 0.3 (MeOH/DCM 5 % v/v); δ_H (400 MHz, DMSO- d_6): 8.07 (s, 2H, exchange with D₂O, SO₂NH₂), 7.83 (s, 2H, exchange with D₂O, NH₂).

Synthesis of phenyl (5-sulfamoyl-1,3,4-thiadiazol-2-yl)carbamate (43)

Phenyl chloroformate (1.5 equiv.) was added dropwise to a solution of **42** (1 g, 1 equiv.) in anhydrous THF (15 mL) at 0 °C under a nitrogen atmosphere, after which the reaction was refluxed for 2 h. The solvent was removed under vacuum, and the crude product was treated with a slush and 12 M HCl until the pH reached 2. The suspension was filtered and washed with H₂O to give a white powder.

Yield 54 %; m.p. 195 °C; silica gel TLC R_f 0.45 (MeOH/DCM 5 % v/v); δ_H (400 MHz, DMSO- d_6): 13.33 (s, 1H, exchange with D₂O, CONH), 8.42 (s, 2H, exchange with D₂O, SO₂NH₂), 7.51 (t, J = 7.8 Hz, 2H, 2 x Ar-H), 7.36 (m, 3H, 3 x Ar-H).

Synthesis of *t*-butyl 4-((5-sulfamoyl-1,3,4-thiadiazol-2-yl)carbamoyl)piperazine-1-carboxylate (44)

DIPEA (cat.) and *t*-butyl piperazine-1-carboxylate (1.2 equiv.) were added to a solution of **43** (0.4 g, 1 equiv.) in anhydrous ACN (5 mL) under a nitrogen atmosphere, after which the resulting suspension was stirred overnight at reflux. The solution was cooled, and a slush was added. Then, the obtained precipitate was filtered to obtain a white powder.

Yield 31 %; m.p. 268–270 °C; silica gel R_f 0.24 (MeOH/DCM 5 % v/v); δ_H (400 MHz, DMSO- d_6): 11.92 (s, 1H, exchange with D₂O, CONH), 8.29 (s, 2H, exchange with D₂O, SO₂NH₂), 3.57 (m, 4H, 2 x CH₂), 3.42 (m, 4H, 2 x CH₂), 1.46 (s, 9H, 3 x CH₃).

General synthesis of intermediates 24, 28, 31, 36, 37, 40 and 45

23, **27**, **30**, **34**, **35**, **39**, and **44** (0.5 g, 1 equiv.) were suspended in 3 M MeOH/HCl or 4 M dioxane/HCl (10 mL) and stirred overnight at RT. The mixtures were evaporated under vacuum, and the resulting powders were suspended in EtOAc or acetone and filtered, yielding white powders.

4-(4-Sulfamoylbenzoyl)piperazin-1-ium chloride (24)

Compound **24** was obtained according to the general procedure previously reported with **23** as the starting material. Yield 86 %; silica gel R_f 0.24 (TFA/MeOH/DCM 4/16/80 % v/v); δ_H (400 MHz, DMSO- d_6): 9.63 (s, 2H, exchange with D₂O, NH₂⁺), 7.93 (d, J = 8.3 Hz, 2H, 2 x Ar-H), 7.69 (d, J = 8.3 Hz, 2H, 2 x Ar-H), 7.55 (s, 2H, exchange with D₂O, SO₂NH₂), 3.87 (s, 2H, CH₂), 3.59 (s, 2H, CH₂), 3.18 (m, 4H, 2 x CH₂).

1-(2-Oxo-2-((4-sulfamoylphenyl)amino)ethyl)piperazine-1,4-diium chloride (28)

Compound **28** was obtained according to the general procedure previously reported with **27** as the starting material. Yield 94 %; silica gel R_f 0.14 (TFA/MeOH/DCM 4/16/80 % v/v); δ_H (400 MHz, DMSO- d_6): 11.23 (s, 1H, exchange with D₂O, NH⁺), 10.50 (s, 1H, exchange with D₂O, CONH), 9.71 (s, 2H, exchange with D₂O, NH₂⁺), 7.85 (m, 4H, 4 x Ar-H), 7.37 (s, 2H, exchange with D₂O, SO₂NH₂), 4.22 (s, 2H, CH₂), 3.49 (m, 8H, 4 x CH₂).

1-(4-Sulfamoylbenzyl)piperazine-1,4-diium chloride (31)

Compound **31** was obtained according to the general procedure previously reported with **30** as the starting material. Yield 91 %; silica gel R_f 0.17 (TFA/MeOH/DCM 4/16/80 % v/v); δ_H (400 MHz, DMSO- d_6): 12.48 (s, 1H, exchange with D₂O, NH⁺), 9.83 (s, 2H, exchange with D₂O, NH₂⁺), 7.90 (m, 4H, 4 x Ar-H), 7.53 (s, 2H, exchange with D₂O, SO₂NH₂), 4.48 (s, 2H, CH₂), 3.49 (m, 8H, 4 x CH₂).

4-((4-Sulfamoylbenzyl)carbamoyl)piperazin-1-ium chloride (36)

Compound **36** was obtained according to the general procedure previously reported with **34** as the starting material. Yield 46 %; silica gel R_f 0.22 (TFA/MeOH/DCM 4/16/80 % v/v); δ_H (400 MHz, DMSO- d_6): 9.32 (s, 2H, exchange with D₂O, NH₂⁺), 7.76 (d, J = 8.8 Hz, 2H, 2 x Ar-H), 7.51 (t, J = 6.0 Hz, 1H, exchange with D₂O, NHCON), 7.44 (d, J = 8.8 Hz, 2H, 2 x Ar-H), 7.34 (s, 2H, exchange with D₂O, SO₂NH₂), 4.30 (d, J = 4.4 Hz, 2H, CH₂), 3.58 (m, 4H, 2 x CH₂), 3.06 (m, 4H, 2 x CH₂).

4-((4-Sulfamoylphenetyl)carbamoyl)piperazin-1-ium chloride (37)

Compound **37** was obtained according to the general procedure previously reported with **35** as the starting material. Yield 68 %; silica gel R_f 0.26 (TFA/MeOH/DCM 4/16/80 % v/v); δ_H (400 MHz, DMSO- d_6): 9.34 (s, 2H, exchange with D₂O, NH₂⁺), 7.76 (d, J = 8.4 Hz, 2H, 2 x Ar-H), 7.39 (d, J = 8.4 Hz, 2H, 2 x Ar-H), 7.33 (s, 2H, exchange with D₂O, SO₂NH₂), 6.96 (m, 1H, exchange with D₂O, NHCON), 3.52 (m, 4H, 2 x CH₂), 3.28 (m, 2H, CH₂), 3.00 (m, 4H, 2 x CH₂), 2.81 (t, J = 6.9 Hz, 2H, CH₂).

4-((4-Sulfamoylphenyl)carbamothioyl)piperazin-1-ium chloride (40)

Compound **40** was obtained according to the general procedure previously reported with **39** as the starting material. Yield 79 %; silica gel R_f 0.31 (TFA/MeOH/DCM 4/16/80 % v/v); δ_H (400 MHz, DMSO- d_6): 9.98 (s, 1H, exchange with D₂O, NHCSN), 9.20 (s, 2H, exchange with D₂O, NH₂⁺), 7.79 (d, J = 8.5 Hz, 2H, 2 x Ar-H), 7.55 (d, J = 8.5 Hz, 2H, 2 x Ar-H), 7.36 (s, 2H, exchange with D₂O, SO₂NH₂), 4.16 (m, 4H, 2 x CH₂), 3.25 (m, 4H, 2 x CH₂).

4-((5-Sulfamoyl-1,3,4-thiadiazol-2-yl)carbamoyl)piperazin-1-ium chloride (45)

Compound **45** was obtained according to the general procedure previously reported with **44** as the starting material. Yield 81 %; silica gel R_f 0.28 (TFA/MeOH/DCM 4/16/80 % v/v); δ_H (400 MHz, DMSO- d_6): 12.02 (s, 1H, exchange with D₂O, CONH), 9.44 (s, 2H, exchange with D₂O, NH₂⁺), 8.35 (s, 2H, exchange with D₂O, SO₂NH₂), 3.83 (m, 4H, 2 x CH₂), 3.19 (m, 4H, 2 x CH₂).

CA inhibitory assay

An Applied Photophysics stopped-flow instrument was used for assaying the CA-catalyzed CO₂ hydration activity [53]. Phenol red (at a concentration of 0.2 mM) was used as an indicator, working at an absorbance maximum of 557 nm, with 20 mM HEPES (pH 7.5) as a buffer and 20 mM Na₂SO₄ (for maintaining constant ionic strength), following the initial rates of the CA-catalyzed CO₂ hydration reaction for a period of 10–100 s. The CO₂ concentrations ranged from 1.7 to 17 mM to determine the kinetic parameters and inhibition constants. For each inhibitor, at least six traces of the initial 5–10 % of the reaction were

used to determine the initial velocity. The uncatalyzed rates were similarly subtracted from the total observed rates. Stock solutions of inhibitor (0.1 mM) were prepared in distilled-deionized water, and dilutions up to 0.01 nM were made thereafter with the assay buffer. The inhibitor and enzyme solutions were preincubated together for 45 min at room temperature prior to the assay to allow for the formation of the E-I complex. The inhibition constants were obtained by nonlinear least-squares methods using PRISM 3 and the Cheng-Prusoff equation, as reported earlier [54,55], which represent the means from at least three different determinations. The enzyme concentrations were in the range of 5–16 nM. All CA isoforms were recombinant and obtained in-house, as reported previously [54–56].

Computational Methodology

ML QSAR model dataset acquisition

The ChEMBL database was used to collect data on compounds that were effective against targets ChEMBL5631, ChEMBL6068, and ChEMBL5767 [57]. The database contains compounds with antibacterial activity; therefore, the initial dataset produced by K_i was chosen; on the other hand, compounds without a known K_i were excluded. After that, 147, 44, and 107 compounds for Mtb β -CA1, Mtb β -CA2, and Mtb β -CA3, respectively, were obtained [57]. All the ML models were built with the KNIME analysis platform [58] using H2O ML [59] nodes and KNIME ML nodes. The molecular descriptors were generated using PaDel [60] standalone software.

Structure normalization and calculation of molecular descriptors

After removing the inconsistencies and biases in the dataset structure, normalization was used to reduce the structural artifacts that may come with initial datasets. Normalization included de-salting, aromatizing, and removing hydrogen. The normalized smiles are then used to generate three-dimensional structures, where the minimization was performed with the UFF forcefield for 1000 iterations. Standalone PaDel software was used to generate the 1D-2D, 3D, and Fingerprints, in which 1D-2D descriptors included Petitjean Number, RingCounts, TPSA, WALKcount, etc.; 3D descriptors included Autocorrelation3D, CPSA, GravitationalIndex, WHIM, etc.; and for the fingerprint, descriptors, PubChem Fingerprinter was used.

Dataset preprocessing and dimensionality reduction

The molecular descriptors were subjected to the “low variance filter” (<0.001) and “correlation filter” (>0.95) nodes for dimensionality reduction; afterward, the dataset was normalized to make it comparable, such that it reduced the biases originating from the magnitudes of the descriptors. Finally, any detected outliers were removed using a “numeric outlier node.”

Genetic algorithm-based feature selection

The feature selection process was based on a genetic algorithm, with a feature selection loop constructed using Naïve Bayes as the learning algorithm. The loop parameters included a population size of 20 and a maximum of 10 generations. For the mechanistic interpretation model, partitioning was performed by stratifying the target variables into Mtb β -CA1, Mtb β -CA2, and Mtb β -CA3. In the individual ML regression QSAR model, stratification was applied to labels as follows: active ($pK_i >6$), intermediate ($pK_i = 5-6$), and inactive ($pK_i <5$).

ML QSAR models for mechanistic interpretation

A mechanistic analysis was conducted to correlate a set of essential chemical features with the experimental inhibition constant values (K_i) across the three target spaces: Mtb β -CA1, Mtb β -CA2, and Mtb β -CA3. A multinomial classification model was constructed for this task, where each molecule was labeled based on the target class with the highest pK_i value. For example, if the pK_i value for Mtb β -CA3 was the highest among the three targets for a particular molecule, it was labeled Mtb

β -CA3. The generated matrix was then coupled with PaDel descriptors and pK_i values of the individual target spaces as the descriptors to develop the model. To construct classification ML models, four types of learners were tested, i.e., naïve Bayes (NB), random forest (RF), gradient boosting (GB), and gradient linear (GL) models. To prevent any chances of overfitting in the models, shallow trees were used (tree = 5 and number of models = 50) for the RF and GB learners. The model performance was evaluated based on matrices such as recall, precision, sensitivity, specificity, F-measure, accuracy, and Cohen's kappa values.

ML-QSAR models

Regression models were also built to predict the pK_i values of the chemical spaces of the Mtb β -CAs. The descriptor generation, dataset curation, dataset preprocessing, and feature selection methods are described earlier (sections 5.4.1-5.4.4). After feature selection, the dataset was partitioned based on the computed k-means cluster in a stratified manner. Here, we tested four types of regression learner nodes: XGBoost (XGB), random forest (RF), gradient boosting (GB), and gradient linear (GL) models. The generated regression models were evaluated based on the mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and R^2 values.

Validation of ML models

All the ML models described herein were validated via the tenfold cross-validation method. This method randomly partitioned the samples into test and training sets, and model performance was evaluated ten times. This method ensures the robustness of the model across the varying distributions of the training dataset.

Applicability domains of the regression models

The applicability domains of the built machine learning models (mechanistic interpretation and ML-QSAR models) were assessed via the chemical space coverage between the training set, test set, and designed query molecules. The chemical spaces were visualized by applying dimensionality reduction techniques, specifically a principal component analysis (PCA) bounding box, to molecular descriptors and fingerprints. The results were then plotted in both 2D and 3D.

ADMET and molecular docking analysis

In silico ADMET profiles of the synthesized compounds were evaluated using SwissADME 9 <http://www.swissadme.ch/> [61] and Osiris Property Explorer software (<https://www.organic-chemistry.org/pr/org/peo/>) [62]. For the molecular docking analysis, experimental crystal structures were available for Mtb β -CA1 (PDB ID: 4YF4, Resolution: 1.80 Å) and Mtb β -CA2 (PDB ID: 2A5V, Resolution: 2.20 Å). However, there is no experimentally determined structure for Mtb β -CA3. Consequently, we modeled its structure using the AlphaFold3 server (<https://alphafoldserver.com>) [63], utilizing the fasta sequence from the UniProt accession ID: P96878, and included the zinc ion in the model. The quality of the resulting protein structure was assessed using positional error heatmaps, Ramachandran plots, and QMEAN score plots. These metrics, including the Ramachandran plot, MolProbity score, and QMEAN Z-score, are essential for evaluating homology models [64]. The Ramachandran plot shows the distribution of residues in allowed phi (ϕ) and psi (ψ) angles, with favored regions indicating stable conformations [64]. The MolProbity score, a composite structural quality indicator, is better when lower, with scores in the 98th percentile or higher considered excellent. The QMEAN Z-score, which compares a model's QMEAN score to high-resolution experimental structures, is optimal when close to 0 or less than 1, indicating good quality [65]. These metrics provide a comprehensive assessment of a protein model's geometry and overall structural quality, making them invaluable tools for validating homology models and ensuring their reliability for further applications.

All the protein structures were preprocessed by stripping off all heteroatoms except the zinc ion. Missing side chains were fixed using Modeller software [66], and energy minimization was performed using

100 steps of steepest descent followed by 10 steps of the conjugate gradient method to remove any structural artifacts. Autodock Vina Zn [67] was used to dock the synthesized compounds 1–13 into the active sites of the zinc-containing metalloenzymes Mtb β -CA1–3. Three-dimensional structures were generated for compounds 1–13 and were subjected to MM2 forcefield optimization. The grid centers were determined based on the geometric coordinates of the zinc ion: Mtb β -CA1 (11.52, −37.56, −17.79), Mtb β -CA2 (11.76, −40.98, −18.336), and Mtb β -CA3 (11.71, −37.55, −16.69) with a grid size of 40x40x40 for each. Binding poses and 2D interaction plots were visualized using Pymol and Ligplot Plus software.

Systems biology approach to identify the probable mechanism of action

Systems biology provides a holistic understanding of biological systems, aiding in identifying mechanisms of action by elucidating complex interactions between molecules and biological pathways. In this study, molecules were selected based on their total average prediction ratios, which were calculated as the average of the prediction ratios for Mtb β -CA1, Mtb β -CA2, and Mtb β -CA3 (equation1).

Total Average Prediction = Prediction ratio of Mtb β – CA1

+ Prediction ratio of Mtb β – CA2 + Prediction ratio of Mtb β – CA3

where the prediction ratio is defined as (equation1):

$$\text{Prediction ratio} = \frac{\text{Experimental pKi}}{\text{Predicted pKi}}$$

Target predictions and disease-related gene mining

After the selection of compounds system analysis, their targets were predicted via an assembly of four tools, namely, Swiss Target Prediction [68], the SEA server [69], the poly-pharmacology server [70] and the SuperPred server [71]. The UniProt server standardized all the predicted target gene names [72]. Incorporating multiple tools with different methodologies has introduced a diverse pool of predicted targets for synthesized compounds. The disease-related targets were mined from three independent databases, namely, DisGeNET [73], OMIM [74], and CTD DB [75], by searching the terms “TB” or “Tuberculosis” in the respective databases. Furthermore, intersecting genes between the compound target gene pool and the TB-related gene pool were identified via the Venny2.0 [76] web server (<https://bioinfogp.cnb.csic.es/tools/venny/index2.0.2.html>).

Protein-protein interactions and hub gene identification

The extracted common genes between the compound target gene pool and TB-related gene pool were subjected to StringDB [77], where the organism was selected as “*Homo sapiens*” to generate a protein-protein interaction (PPI) network of the common genes. The comprehensive dense PPI network was exported to the Cytoscape platform, and the cytoHubba plugin was used to find the top ten hub genes based on the degree method.

Gene enrichment, gene ontology, KEGG pathway analysis, and tissue-specific localization

The identified hub genes were subjected to gene enrichment analysis via the DAVID [78] webserver. The list of the top ten hub genes was submitted to the server, in which the “Official gene symbol” was selected as the identifier, “*Homo sapiens*” was selected as the choice of the organism, and the list type was selected as the “Gene list.” In the KEGG pathway [79] analysis, terms not relevant to TB or other infectious diseases were discarded, and the final pool of terms was filtered with a p-value > 0.05 threshold. Gene ontologies [80] were also performed based on biological processes (BP), cellular compartment (CC), and molecular functions (MF), which were filtered with a p-value > 0.05. The results are depicted as gene enrichment bubble plots constructed using the SRPLOTs [81] webserver <https://www.bioinformatics.com.cn/en>.

cn/en.

To assess potential side effects associated with synthesized compounds (9, 8, 10, 12, 13, and AAZ), we utilized NetInfer, a network-based inference server known for predicting adverse drug events (ADEs) and chemical side effects (PMID: 32687354). The predictions were aggregated and analyzed using hierarchical clustering to identify clusters of similar side effects. From these predictions, the top 20 terms with high frequency were extracted. Each term was classified using ChatGPT to better understand its nature and potential implications.

Additionally, organ-specific localization of the hub genes was explored via single-cell expression analysis via the GTEX [50] server <https://gtexportal.org/home/multiGeneSingleCellQueryPage>.

Compound target interactions network construction

To thoroughly understand the complex interactions between the compounds (layer 1), TB-related genes (layer 2), and enriched KEGG pathways (layer 3), a tripartite network was constructed on the Cytoscape [82] platform. Subsequently, to understand the interaction networks in the context of tuberculosis, a hierarchical network was constructed to determine the influence of a compound with a specific gene that is ultimately related to this disease.

Hierarchical clustering of compound target genes and their associated genes, pathways, phenotypes, and potential side effects

We calculated Jaccard similarity coefficients, plotted as heatmaps to understand the similarities between the synthesized molecules in terms of their genes, physical descriptors, phenotypes, and pathways. The Jaccard similarity coefficient is defined as the size of the intersection divided by the size of one of the sets, and mathematically, it is expressed as

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

A and B represent the two sets for which similarity coefficients are calculated.

First, a biological network was built, incorporating chemicals with their predicted target genes, pathways, and phenotypes, using data gathered from the CTD database [75]. We also included chemicals and their probable side effects using the Netinfer tool [83]. to better understand our designed inhibitors' potential side effects.

After constructing the biological network, we performed Jaccard index calculations for synthesized chemicals based on common genes, pathways, phenotypes, and side effects. By comparing the overlap of elements within different genes, pathways, phenotypes, and side effects, we quantified the degree of similarity between these entities. This approach provides a robust measure of how closely related chemicals are in the context of their targeted genes, pathways, phenotypes, and side effects, facilitating a deeper understanding of their interconnections.

Once the Jaccard index scores were obtained, hierarchical clustering was applied. Hierarchical clustering is a method of cluster analysis that builds a hierarchy of clusters. In this context, similar genes, pathways, phenotypes, and side effects, chemicals were grouped based on their Jaccard index scores, allowing us to identify closely related groups and gain insights into their relationships. To visualize the results, clustergrams were produced to show the hierarchical clusters, and word clouds were generated to highlight common terms for genes, pathways, phenotypes, and side effects. These visualizations helped interpret the data and understand the connections between different biological entities.

CRediT authorship contribution statement

Simone Giovannuzzi: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sagar Singh Shyamal:** Writing –

review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ratul Bhowmik**: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Rajarshi Ray**: Writing – review & editing, Supervision, Formal analysis, Validation, Visualization. **Ajay Manaihiya**: Writing – review & editing, Visualization, Supervision, Formal analysis, Resources, Validation. **Fabrizio Carta**: Writing – review & editing, Supervision, Formal analysis. **Seppo Parkkila**: Writing – review & editing, Supervision, Funding acquisition, Formal analysis. **Ashok Aspatwar**: Writing – review & editing, Supervision, Funding acquisition, Formal analysis. **Claudiu T. Supuran**: Writing – review & editing, Supervision, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors extend their sincere appreciation to the following organizations for their invaluable support and funding: 1) the Finnish Cultural Foundation; 2) the Tampere Tuberculosis Foundation (Ashok Aspatwar); 3) the Academy of Finland; and 4) the Jane and Aatos Erkko Foundation (Seppo Parkkila).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.compbiomed.2024.109029>.

References

- [1] P. Glaziou, K. Floyd, M.C. Raviglione, Global epidemiology of tuberculosis, *Semin. Respir. Crit. Care Med.* 39 (2018) 271–285, <https://doi.org/10.1055/s-0038-1651492>.
- [2] S. Who, Geneva, WHO, Global Tuberculosis Report (2021) 1–57.
- [3] G.F.S. Fernandes, A.M. Thompson, D. Castagnolo, W.A. Denny, J.L. Dos Santos, Tuberculosis drug discovery: challenges and new horizons, *J. Med. Chem.* 65 (2022) 7489–7531, <https://doi.org/10.1021/acs.jmedchem.2c00227>.
- [4] L. Zhuang, Z. Ye, L. Li, L. Yang, W. Gong, Next-generation TB vaccines: progress, challenges, and prospects, *Vaccines* 11 (2023), <https://doi.org/10.3390/vaccines11081304>.
- [5] S.T. Cole, Learning from the genome sequence of *Mycobacterium tuberculosis* H37Rv, *FEBS Lett.* 452 (1999) 7–10, [https://doi.org/10.1016/S0014-5793\(99\)00536-0](https://doi.org/10.1016/S0014-5793(99)00536-0).
- [6] B. Cao, X. Mijiti, L. Le Deng, Q. Wang, J.J. Yu, A. Anwaierjiang, C. Qian, M. Li, D. A. Fang, Y. Jiang, L.L. Zhao, X. Zhao, K. Wan, H. Liu, G. Li, X. Yuan, Genetic characterization conferred co-resistance to isoniazid and ethionamide in mycobacterium tuberculosis isolates from Southern Xinjiang, China, *Infect. Drug Resist.* 16 (2023) 3117–3135, <https://doi.org/10.2147/IDR.S407525>.
- [7] C.L. Daley, J.A. Caminero, Management of multidrug-resistant tuberculosis, *Semin. Respir. Crit. Care Med.* 39 (2018) 310–324, <https://doi.org/10.1055/s-0038-1661383>.
- [8] A.K. Saxena, A. Singh, Mycobacterial tuberculosis enzyme targets and their inhibitors, *Curr. Top. Med. Chem.* 19 (2019) 337–355, <https://doi.org/10.2174/1568026619666190219105722>.
- [9] G. Kaul, E. Kapoor, A. Dasgupta, S. Chopra, Management of multidrug-resistant tuberculosis in the 21st century, *Drugs Today* 55 (2019) 215–224, <https://doi.org/10.1358/DOT.2019.55.3.2927587>.
- [10] I. Nishimori, T. Minakuchi, A. Maresca, F. Cartab, A. Scozzafava, C.T. Supuran, The β -carbonic anhydrases from *Mycobacterium tuberculosis* as drug targets, *Curr. Pharmaceut. Des.* 16 (2010) 3300–3309, <https://doi.org/10.2174/138161210793429814>.
- [11] A. Aspatwar, V. Kairys, S. Rala, M. Parikka, M. Bozdag, F. Carta, C.T. Supuran, S. Parkkila, Mycobacterium tuberculosis β -carbonic anhydrases: novel targets for developing antituberculosis drugs, *Int. J. Mol. Sci.* 20 (2019), <https://doi.org/10.3390/IJMS20205153>.
- [12] A. Aspatwar, M. Hammar En A Å, S. Koskinen, B. Luukinen, H. Barker, F. Carta, C. T. Supuran, M. Parikka, S. Parkkila, b-CA-specific inhibitor dithiocarbamate Fc14-584B: a novel antimycobacterial agent with potential to treat drug-resistant tuberculosis, (n.d.). <https://doi.org/10.1080/14756366.2017.1332056>.
- [13] I. Nishimori, T. Minakuchi, T. Kohsaki, S. Onishi, H. Takeuchi, D. Vullo, A. Scozzafava, C.T. Supuran, Carbonic anhydrase inhibitors: the beta-carbonic anhydrase from *Helicobacter pylori* is a new target for sulfonamide and sulfamate inhibitors, *Bioorg. Med. Chem. Lett.* 17 (2007) 3585–3594, <https://doi.org/10.1016/j.bmcl.2007.04.063>.
- [14] A. Innocenti, F.A. Mühlischlegel, R.A. Hall, C. Steegborn, A. Scozzafava, C. T. Supuran, Carbonic anhydrase inhibitors: inhibition of the b-class enzymes from the fungal pathogens *Candida albicans* and *Cryptococcus neoformans* with simple anions, <https://doi.org/10.1016/j.bmcl.2008.07.122>, 2008.
- [15] C.T. Supuran, Structure-based drug discovery of carbonic anhydrase inhibitors, *J. Enzym. Inhib. Med. Chem.* 27 (2012) 759–772, <https://doi.org/10.3109/14756366.2012.672983>.
- [16] E.V. Koonin, K.S. Makarova, Origins and evolution of CRISPR-Cas systems, *Phil. Trans. Biol. Sci.* 374 (2019) 20180087, <https://doi.org/10.1098/rstb.2018.0087>.
- [17] C.T. Supuran, How many carbonic anhydrase inhibition mechanisms exist? *J. Enzym. Inhib. Med. Chem.* 31 (2016) 345–360, <https://doi.org/10.3109/14756366.2015.1122001>.
- [18] F. Carta, C.T. Supuran, Diuretics with carbonic anhydrase inhibitory action: a patent and literature review (2005 – 2013), *Expert Opin. Ther. Pat.* 23 (2013) 681–691, <https://doi.org/10.1517/13543776.2013.780598>.
- [19] C.T. Supuran, Carbonic anhydrase inhibition and the management of neuropathic pain, *Expert Rev. Neurother.* 16 (2016) 961–968, <https://doi.org/10.1080/14737175.2016.1193009>.
- [20] C. Capasso, C.T. Supuran, An overview of the alpha-, beta- and gamma-carbonic anhydrases from *Bacteria*: can bacterial carbonic anhydrases shed new light on evolution of bacteria? *J. Enzym. Inhib. Med. Chem.* 30 (2015) 325–332, <https://doi.org/10.3109/14756366.2014.910202>.
- [21] S. Isik, F. Kockar, M. Aydin, O. Arslan, O.O. Guler, A. Innocenti, A. Scozzafava, C. T. Supuran, Carbonic anhydrase inhibitors: inhibition of the beta-class enzyme from the yeast *Saccharomyces cerevisiae* with sulfonamides and sulfamates, *Bioorg. Med. Chem.* 17 (2009) 1158–1163, <https://doi.org/10.1016/J.BMC.2008.12.035>.
- [22] A. Nocentini, S.M. Osman, I.A. Almeida, V. Cardoso, F.A.S. Alasmay, Z. AlOthman, A.B. Vermelho, P. Gratter, C.T. Supuran, Appraisal of anti-protozoan activity of nitroaromatic benzenesulfonamides inhibiting carbonic anhydrases from *Trypanosoma cruzi* and *Leishmania donovani*, *J. Enzym. Inhib. Med. Chem.* 34 (2019) 1164–1171, <https://doi.org/10.1080/14756366.2019.1626375>.
- [23] S. Bua, S.M. Osman, S. Del Prete, C. Capasso, Z. AlOthman, A. Nocentini, C. T. Supuran, Click-tailed benzenesulfonamides as potent bacterial carbonic anhydrase inhibitors for targeting *Mycobacterium tuberculosis* and *Vibrio cholerae*, *Bioorg. Chem.* 86 (2019) 183–186, <https://doi.org/10.1016/J.BIOORG.2019.01.065>.
- [24] A. Aspatwar, J.Y. Winum, F. Carta, C.T. Supuran, M. Hammaren, M. Parikka, S. Parkkila, Carbonic anhydrase inhibitors as novel drugs against mycobacterial β -carbonic anhydrases: an update on in vitro and in vivo studies, *Molecules* 23 (2018), <https://doi.org/10.3390/MOLECULES23112911>.
- [25] M. Minakuchi, I. Nishimori, D. Vullo, A. Scozzafava, C.T. Supuran, Molecular cloning, characterization, and inhibition studies of the Rv1284 beta-carbonic anhydrase from *Mycobacterium tuberculosis* with sulfonamides and a sulfamate, *J. Med. Chem.* 52 (2009) 2226–2232, <https://doi.org/10.1021/JM9000488>.
- [26] I. Nishimori, T. Minakuchi, D. Vullo, A. Scozzafava, A. Innocenti, C.T. Supuran, Carbonic anhydrase inhibitors. Cloning, characterization, and inhibition studies of a new beta-carbonic anhydrase from *Mycobacterium tuberculosis*, *J. Med. Chem.* 52 (2009) 3116–3120, <https://doi.org/10.1021/JM9003126>.
- [27] C. Marzano, L. Ronconi, F. Chiara, M.C. Giron, I. Faustinielli, P. Cristofori, A. Trevisan, D. Fregona, Gold(III)-dithiocarbamate anticancer agents: activity, toxicology and histopathological studies in rodents, *Int. J. Cancer* 129 (2011) 487–496, <https://doi.org/10.1002/IJC.25684>.
- [28] A. Maresca, F. Carta, D. Vullo, C.T. Supuran, Dithiocarbamates strongly inhibit the β -class carbonic anhydrases from *Mycobacterium tuberculosis*, *J. Enzym. Inhib. Med. Chem.* 28 (2013) 407–411, <https://doi.org/10.3109/14756366.2011.641015>.
- [29] L.H. Hall, L.B. Kier, Electrotological state indices for atom types: a novel combination of electronic, topological, and valence state information, *J. Chem. Inf. Comput. Sci.* 35 (1995) 1039–1045, <https://doi.org/10.1021/C100028A014/ASSET/C100028A014.FP.PNG.V03>.
- [30] A. Tropsha, A. Golbraikh, Predictive QSAR modeling workflow, model applicability domains, and virtual screening, *Curr. Pharmaceut. Des.* 13 (2007) 3494–3504, <https://doi.org/10.2174/138161207782794257>.
- [31] A. Cherkasov, E.N. Muratov, D. Fourches, A. Varnek, I.I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y.C. Martin, R. Todeschini, V. Consonni, V.E. Kuz'Min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, A. Tropsha, QSAR modeling: where have you been? Where are you going to? *J. Med. Chem.* 57 (2014) 4977–5010, https://doi.org/10.1021/JM4004285/ASSET/IMAGES/LARGE/JM-2013-004285_0008.JPEG.
- [32] I. Sushko, S. Novotarskyi, R. Körner, A.K. Pandey, A. Cherkasov, J. Li, P. Gramatica, K. Hansen, T. Schroeter, K.R. Müller, L. Xi, H. Liu, X. Yao, T. Öberg, F. Hormozdiari, P. Dao, C. Sahinalp, R. Todeschini, P. Polishchuk, A. Artemenko, V. Kuz'Min, T.M. Martin, D.M. Young, D. Fourches, E. Muratov, A. Tropsha, I. Baskin, D. Horvath, G. Marcou, C. Muller, A. Varnek, V.V. Prokopenko, I. V. Tetko, Applicability domains for classification problems: benchmarking of distance to models for ames mutagenicity set, *J. Chem. Inf. Model.* 50 (2010) 2094–2111, <https://doi.org/10.1021/ci100253r>.

- [33] J. Veiga-Matos, A.I. Morales, M. Prieto, F. Remião, R. Silva, Study models of drug–drug interactions involving P-glycoprotein: the potential benefit of P-glycoprotein modulation at the kidney and intestinal levels, *Molecules* 28 (2023) 7532, <https://doi.org/10.3390/molecules28227532>.
- [34] R. Callaghan, F. Luk, M. Bebaawy, Inhibition of the multidrug resistance P-glycoprotein: time for a change of strategy? *Drug Metabol. Dispos.* 42 (2014) 623–631, <https://doi.org/10.1124/dmd.113.056176>.
- [35] X. Wen, J.-S. Wang, P.J. Neuvonen, J.T. Backman, Isoniazid is a mechanism-based inhibitor of cytochrome P 450 1A2, 2A6, 2C19 and 3A4 isoforms in human liver microsomes, *Eur. J. Clin. Pharmacol.* 57 (2002) 799–804, <https://doi.org/10.1007/s00228-001-0396-3>.
- [36] S. Bhatt, S. Dhiman, V. Kumar, A. Gour, D. Manhas, K. Sharma, P.K. Ojha, N. Nandi, Assessment of the CYP1A2 inhibition-mediated drug interaction potential for pinocembrin using *in silico*, *in vitro*, and *in vivo* approaches, *ACS Omega* 7 (2022) 20321–20331, <https://doi.org/10.1021/acsomega.2c02315>.
- [37] R. Stemkens, V. de Jager, R. Dawson, A.H. Diacon, K. Narunsky, S.D. Padayachee, M.J. Boeree, S.W. van Beek, A. Colbers, M.J.H. Coenen, E.M. Svensson, U. Fuhr, P. P.J. Phillips, L.H.M. te Brake, R.E. Aarnoutse, Drug interaction potential of high-dose rifampicin in patients with pulmonary tuberculosis, *Antimicrob. Agents Chemother.* 67 (2023), <https://doi.org/10.1128/aac.00683-23>.
- [38] O.V. Sobolev, P.V. Afonine, N.W. Moriarty, M.L. Hekkelman, R.P. Joosten, A. Perrakis, P.D. Adams, A global ramachandran score identifies protein structures with unlikely stereochemistry, *Structure* 28 (2020) 1249–1258.e2, <https://doi.org/10.1016/j.str.2020.08.005>.
- [39] D. Coskun, M. Lihan, J.P.G.L.M. Rodrigues, M. Vass, D. Robinson, R.A. Friesner, E. B. Miller, Using AlphaFold and experimental structures for the prediction of the structure and binding affinities of GPCR complexes via induced fit docking and free energy perturbation, *J. Chem. Theor. Comput.* 20 (2024) 477–489, <https://doi.org/10.1021/acs.jctc.3c00839>.
- [40] V. Scardino, J.I. Di Filippo, C.N. Cavasotto, How good are AlphaFold models for docking-based virtual screening? *iScience* 26 (2023) 105920 <https://doi.org/10.1016/j.isci.2022.105920>.
- [41] A. Aspatwar, V. Kairys, S. Rala, M. Parikka, M. Bozdag, F. Carta, C.T. Supuran, S. Parkkila, Mycobacterium tuberculosis β -carbonic anhydrases: novel targets for developing antituberculosis drugs, *Int. J. Mol. Sci.* 20 (2019) 5153, <https://doi.org/10.3390/ijms20205153>.
- [42] K.Y. Ji, C. Liu, Z.Q. Liu, Y.F. Deng, T.J. Hou, D.S. Cao, Comprehensive assessment of nine target prediction web services: which should we choose for target fishing? *Briefings Bioinf.* 24 (2023) <https://doi.org/10.1093/bib/bbad014>.
- [43] J.K. Huang, D.E. Carlin, M.K. Yu, W. Zhang, J.F. Kreisberg, P. Tamayo, T. Ideker, Systematic evaluation of molecular networks for discovery of disease genes, *Cell Syst* 6 (2018) 484–495.e5, <https://doi.org/10.1016/j.cels.2018.03.001>.
- [44] Y. Wu, A. Riehle, B. Pollmeier, S. Kadow, F. Schumacher, M. Drab, B. Kleuser, E. Gulbins, H. Grassmé, Caveolin-1 affects early mycobacterial infection and apoptosis in macrophages and mice, *Tuberculosis* 147 (2024), <https://doi.org/10.1016/j.tube.2024.102493>.
- [45] T. Liu, L. Zhang, D. Joo, S.C. Sun, NF- κ B signaling in inflammation, *Signal Transduct. Targeted Ther.* 2 (2017), <https://doi.org/10.1038/sigtrans.2017.23>.
- [46] M.G. Gutierrez, B.B. Mishra, L. Jordao, E. Elliott, E. Anes, G. Griffiths, NF- κ B activation controls phagolysosome fusion-mediated killing of mycobacteria by macrophages, *J. Immunol.* 181 (2008) 2651–2663, <https://doi.org/10.4049/JIMMUNOL.181.4.2651>.
- [47] N. Wang, H. Liang, K. Zen, Molecular mechanisms that influence the macrophage M1-M2 polarization balance, *Front. Immunol.* 5 (2014), <https://doi.org/10.3389/fimmu.2014.00614>.
- [48] M. Wagener, J.C. Hoving, H. Ndlovu, M.J. Marakalala, Dectin-1-Syk-CARD9 signaling pathway in TB immunity, *Front. Immunol.* 9 (2018), <https://doi.org/10.3389/fimmu.2018.00225>.
- [49] I.E. Wartz, K.M. O'Rourke, H. Zhou, M. Eby, L. Aravind, S. Seshagiri, P. Wu, C. Wiesmann, R. Baker, D.L. Boone, A. Ma, E.V. Koonin, V.M. Dixit, De-ubiquitination and ubiquitin ligase domains of A20 downregulate NF- κ B signalling, *Nature* 430 (7000) (2004) 694–699, <https://doi.org/10.1038/nature02794>, 430 (2004).
- [50] V. Gangwar, A. Garg, K. Lomore, K. Korla, S.S. Bhat, R.P. Rao, M. Rafiq, R. Kumawath, B.V. Uddagiri, V.V. Kareenahalli, Immunomodulatory effects of a concretion of natural bioactive compounds—mechanistic insights, *Biomedicines* 9 (2021) 1522, <https://doi.org/10.3390/biomedicines9111522>.
- [51] K.I.R. Notarte, M.T.J. Quimque, I.T. Macaranas, A. Khan, A.M. Pastrana, O. B. Villaflores, H.C.P. Arturo, D.Y.H. Pilapil IV, S.M.M. Tan, D.-Q. Wei, A. Wenzel-Storjohann, D. Tasdemir, C.-H. Yen, S.Y. Ji, G.-Y. Kim, Y.H. Choi, A.P.G. Macabeo, Attenuation of lipopolysaccharide-induced inflammatory responses through inhibition of the NF- κ B pathway and the increased NRF2 level by a flavonol-enriched *n*-butanol fraction from *Uvaria alba*, *ACS Omega* 8 (2023) 5377–5392, <https://doi.org/10.1021/acsomega.2c06451>.
- [52] A. Jenab, R. Roghanian, G. Emtiazi, Bacterial natural compounds with anti-inflammatory and immunomodulatory properties (mini review), *Drug Des. Dev. Ther.* 14 (2020) 3787–3801, <https://doi.org/10.2147/DDDT.S261283>.
- [53] R.G. Kxalifah, The carbon dioxide hydration activity of carbonic anhydrase I. STOP-FLOW kinetic studies on the native human isoenzymes B and C", *J. Biol. Chem.* 246 (1971) 2561–2573, [https://doi.org/10.1016/S0021-9258\(18\)62326-9](https://doi.org/10.1016/S0021-9258(18)62326-9).
- [54] A. Nocentini, A. Bonardi, P. Gratteri, B. Cerra, A. Gioiello, C.T. Supuran, Steroids interfere with human carbonic anhydrase activity by using alternative binding mechanisms, *J. Enzym. Inhib. Med. Chem.* 33 (2018) 1453–1459, <https://doi.org/10.1080/14756366.2018.1512597>.
- [55] A. Nocentini, E. Trallori, S. Singh, C.L. Lomelino, G. Bartolucci, L. Di Cesare Mannelli, C. Ghelardini, R. McKenna, P. Gratteri, C.T. Supuran, 4-Hydroxy-3-nitro-5-ureido-benzenesulfonamides selectively target the tumor-associated carbonic anhydrase isoforms IX and XII showing hypoxia-enhanced antiproliferative profiles, *J. Med. Chem.* 61 (2018) 10860–10874, <https://doi.org/10.1021/ACS.JMEDCHEM.8B01504>.
- [56] C.T. Supuran, Carbon- versus sulphur-based zinc binding groups for carbonic anhydrase inhibitors? *J. Enzym. Inhib. Med. Chem.* 33 (2018) 485–495, <https://doi.org/10.1080/14756366.2018.1428572>.
- [57] B. Zdr azil, E.F. elix, F. Hunt er, E.J. Manners, J. ames Blac ksha, S. Corbett, M. de Veij, H. is Ioannidis, D. Mendez Lopez, J.F. Mosquera, M.P. ia aula Magar inos, N. Bosc, R. do Ar cila, T. Kizilören, A. Gaulton, A.P. atr ícia Bento, M.F. A dasme, P. er Monec ke, G.A. ory Landrum, A.R. Leach, A. Patrícia Bento, The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods, *Nucleic Acids Res.* 2023 (2023) 1–13, <https://doi.org/10.1093/NAR/GKAD1004>.
- [58] M.R. Berthold, N. Cebron, F. Dill, T.R. Gabriel, T. Kötter, T. Meinl, P. Ohl, C. Sieb, K. Thiel, B. Wiswedel, KNIME: The Konstanz Information Miner (2008) 319–326, https://doi.org/10.1007/978-3-540-78246-9_38.
- [59] Erin LeDell, Sebastien Poirier, H2O AutoML: scalable automatic machine learning, in: 7th ICDL Workshop on Automated Machine Learning (AutoML), 2020.
- [60] C.W. Yap, PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints, *J. Comput. Chem.* 32 (2011) 1466–1474, <https://doi.org/10.1002/JCC.21707>.
- [61] A. Daina, O. Michielin, V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, *Sci. Rep.* 7 (1) (2017) 1–13, <https://doi.org/10.1038/srep42717>, 7 (2017).
- [62] T. Sander, OSIRIS property explorer. *Organic Chemistry Portal*, 2001.
- [63] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A.J. Ballard, J. Bambrick, S.W. Bodenstein, D.A. Evans, C.-C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A.I. Cowen-Rivers, A. Cowie, M. Figurnov, F.B. Fuchs, H. Gladman, R. Jain, Y.A. Khan, C.M. R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E.D. Zhong, M. Zielinski, A. Židek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis, J.M. Jumper, Accurate structure prediction of biomolecular interactions with AlphaFold 3, *Nature* 630 (2024) 493–500, <https://doi.org/10.1038/s41586-024-07487-w>.
- [64] S.A. Hollingsworth, P.A. Karplus, A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins, *Biomol. Concepts* 1 (2010) 271–283, <https://doi.org/10.1515/bmc.2010.022>.
- [65] V.B. Chen, W.B. Arendall, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L. W. Murray, J.S. Richardson, D.C. Richardson, *MolProbity*: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr D Biol Crystallogr* 66 (2010) 12–21, <https://doi.org/10.1107/S09074449090042073>.
- [66] N. Eswar, B. Webb, M.A. Marti-Renom, M.S. Madhusudan, D. Eramian, M. Shen, U. Pieper, A. Sali, Comparative protein structure modeling using modeller, in: *Current Protocols in Bioinformatics*/Editorial Board, Andreas D. Baveanis ... [et Al.], vol. 5, 2006, <https://doi.org/10.1002/0471250953.BI0506S15>. Unit.
- [67] J. Eberhardt, D. Santos-Martins, A.F. Tillack, S. Forli, AutoDock vina 1.2.0: new docking methods, expanded force field, and Python bindings, *J. Chem. Inf. Model.* 61 (2021) 3891–3898, <https://doi.org/10.1021/acs.jcim.1c00203>.
- [68] A. Daina, O. Michielin, V. Zoete, SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules, *Nucleic Acids Res.* 47 (2019) W357–W364, <https://doi.org/10.1093/NAR/GKZ382>.
- [69] M.J. Keiser, B.L. Roth, B.N. Armbruster, P. Ernsberger, J.J. Irwin, B.K. Shoichet, Relating protein pharmacology by ligand chemistry, *Nat. Biotechnol.* 25 (2007) 197–206, <https://doi.org/10.1038/NBT1284>.
- [70] M. Awale, J.L. Reymond, Polypharmacology browser PPB2: target prediction combining nearest neighbors with machine learning, *J. Chem. Inf. Model.* 59 (2019) 10–17, <https://doi.org/10.1021/acs.jcim.8b00524>.
- [71] K. Gallo, A. Goede, R. Preissner, B.-O. Gohlke, SuperPred 3.0: drug classification and target prediction—a machine learning approach, *Nucleic Acids Res.* 50 (2022) W726–W731, <https://doi.org/10.1093/nar/gkac297>.
- [72] T. UniProt Consortium, UniProt: the universal protein knowledgebase, *Nucleic Acids Res.* 46 (2018) 2699, <https://doi.org/10.1093/nar/gky092>, 2699.
- [73] J. Piñero, J.M. Ramírez-Anguila, J. Saüch-Pitarch, F. Ronzano, E. Centeno, F. Sanz, L.I. Furlong, The DisGeNET knowledge platform for disease genomics: 2019 update, *Nucleic Acids Res.* 48 (2020) D845–D855, <https://doi.org/10.1093/NAR/GKZ1021>.
- [74] J.S. Amberger, C.A. Bocchini, A.F. Scott, A. Hamosh, OMIM.org: leveraging knowledge across phenotype–gene relationships, *Nucleic Acids Res.* 47 (2019) D1038–D1043, <https://doi.org/10.1093/NAR/GKY1151>.
- [75] A.P. Davis, T.C. Wiegiers, R.J. Johnson, D. Sciaky, J. Wiegiers, C.J. Mattingly, Comparative toxicogenomics database (CTD): update 2023, *Nucleic Acids Res.* 51 (2022), <https://doi.org/10.1093/nar/gkac833>.
- [76] J.C. Oliveros, An Interactive Tool for Comparing Lists with Venn's Diagrams, 2015. <http://Bioinfofp.Cnb.Csic.Es/Tools/Venny/Index.Html>.
- [77] D. Szklarczyk, A.L. Gable, K.C. Nastou, D. Lyon, R. Kirsch, S. Pyysalo, N. T. Doncheva, M. Legeay, T. Fang, P. Bork, L.J. Jensen, C. von Mering, The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets, *Nucleic Acids Res.* 49 (2021) D605–D612, <https://doi.org/10.1093/NAR/GKAA1074>.
- [78] B.T. Sherman, M. Hao, J. Qiu, X. Jiao, M.W. Baseler, H.C. Lane, T. Imamichi, W. Chang, DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update), *Nucleic Acids Res.* 50 (2022) W216–W221, <https://doi.org/10.1093/nar/gkac194>.

- [79] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, *Nucleic Acids Res.* 28 (2000) 27–30, <https://doi.org/10.1093/NAR/28.1.27>.
- [80] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Gene Ontology: tool for the unification of biology, *Nat. Genet.* 25 (2000) 25–29, <https://doi.org/10.1038/75556>.
- [81] D. Tang, M. Chen, X. Huang, G. Zhang, L. Zeng, G. Zhang, S. Wu, Y. Wang, SRplot: a free online platform for data visualization and graphing, *PLoS One* 18 (2023), <https://doi.org/10.1371/journal.pone.0294236>.
- [82] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res.* 13 (2003) 2498–2504, <https://doi.org/10.1101/GR.1239303>.
- [83] Z. Wu, Y. Peng, Z. Yu, W. Li, G. Liu, Y. Tang, NetInfer: a web server for prediction of targets and therapeutic and adverse effects via network-based inference methods, *J. Chem. Inf. Model.* 60 (2020) 3687–3691, <https://doi.org/10.1021/acs.jcim.0c00291>.