

Article

Computational Drug Repurposing for Antituberculosis Therapy: Discovery of Multi-Strain Inhibitors

Valeria V. Kleandrova ¹, Marcus T. Scotti ² and Alejandro Speck-Planche ^{2,*}

¹ Laboratory of Fundamental and Applied Research of Quality and Technology of Food Production, Moscow State University of Food Production, Volokolamskoe shosse 11, 125080 Moscow, Russia; valeria.kleandrova@gmail.com

² Postgraduate Program in Natural and Synthetic Bioactive Products, Federal University of Paraíba, João Pessoa 58051-900, Brazil; mtscotti@gmail.com

* Correspondence: alejspivanovich@gmail.com; Tel.: +55-7-9262539385

Abstract: Tuberculosis remains the most afflicting infectious disease known by humankind, with one quarter of the population estimated to have it in the latent state. Discovering antituberculosis drugs is a challenging, complex, expensive, and time-consuming task. To overcome the substantial costs and accelerate drug discovery and development, drug repurposing has emerged as an attractive alternative to find new applications for “old” drugs and where computational approaches play an essential role by filtering the chemical space. This work reports the first multi-condition model based on quantitative structure–activity relationships and an ensemble of neural networks (mtc-QSAR-EL) for the virtual screening of potential antituberculosis agents able to act as multi-strain inhibitors. The mtc-QSAR-EL model exhibited an accuracy higher than 85%. A physicochemical and fragment-based structural interpretation of this model was provided, and a large dataset of agency-regulated chemicals was virtually screened, with the mtc-QSAR-EL model identifying already proven antituberculosis drugs while proposing chemicals with great potential to be experimentally repurposed as antituberculosis (multi-strain inhibitors) agents. Some of the most promising molecules identified by the mtc-QSAR-EL model as antituberculosis agents were also confirmed by another computational approach, supporting the capabilities of the mtc-QSAR-EL model as an efficient tool for computational drug repurposing.

Keywords: artificial neural networks; drug repurposing; ensemble; MLP; QSAR; tuberculosis; strain; virtual screening



Citation: Kleandrova, V.V.; Scotti, M.T.; Speck-Planche, A. Computational Drug Repurposing for Antituberculosis Therapy: Discovery of Multi-Strain Inhibitors. *Antibiotics* **2021**, *10*, 1005. <https://doi.org/10.3390/antibiotics10081005>

Academic Editor: Alessia Carocci

Received: 1 August 2021

Accepted: 17 August 2021

Published: 19 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Tuberculosis (TB) constitutes the deadliest infectious disease that afflicts humankind. The causative agent of TB, *Mycobacterium tuberculosis* (*Mtb*), was responsible in 2018 for causing more than 10 million cases of active TB, resulting in 1.5 million deaths [1]. Despite the efforts of the scientific community in providing anti-TB therapies through the process known as drug discovery, several aspects pose great challenges for the safe and successful eradication of TB. On one hand, the thick cell wall formed by mycolic acids, the ability of certain enzymes to modify/inactivate drug molecules, the presence of drug efflux systems, and the occurrence of spontaneous mutations in the bacterial genome are the main biological attributes that make *Mtb* considerably resistant (multidrug-resistant TB) to current anti-TB treatments [2]. On the other hand, the FDA-approved anti-TB drugs are associated with a low diversity of mechanisms of action [3] and exhibit a wide range of side effects [4]. All these ideas, together with the complexity and remarkable expenditure of time and financial resources during the drug discovery process [5], demonstrate that TB is hard to treat and that efficacious anti-TB agents are urgently needed.

A plausible way to tackle the TB problem is to apply the drug repurposing philosophy, which focuses on finding new applications for “old” drugs [6]. In this sense, in the context

of identifying novel anti-TB agents, computational approaches have played an essential role by filtering the chemical space, providing different tools that can speed up the discovery of molecules able to inhibit the growth of *Mtb* [7–14]. However, all the computational approaches reported to date have at least one of the following aspects that prevent their full exploitation in virtual screening scenarios: (a) the anti-TB activity is predicted in a generic manner or by considering either only one target protein or *Mtb* strain, (b) small datasets of structurally related chemicals are used to build the computational models, (c) there is no information regarding the experimental protocol or the assay time employed to assess the inhibitory activity against *Mtb*, and (d) no insights are provided concerning the physicochemical properties and structural features that are required in a molecule to have anti-TB activity.

To solve the aforementioned limitations, several researchers have emphasized the use of interpretable *in silico* models focused on a combination of perturbation theory concepts and machine learning techniques (PTML) [15–17], which can integrate different sources of chemical and biological data, enabling the simultaneous prediction of multiple biological endpoints against many targets of varying degrees of complexity. Seminal works on PTML models have found successful applications in diverse research areas such as infectious diseases [18,19], oncology [20,21], neuroscience [22–25], proteomics [26], metabolomics [27], nanotechnology [28–31], toxicology [32], and immunology and immunotoxicity [33,34].

Bearing in mind all the previous ideas, we report in this work a special case of the PTML modeling methodology in the context of the search for novel anti-TB chemical therapies, establishing the theoretical foundations for the computational repurposing of drugs as anti-TB agents. Thus, we have developed here a multi-condition model based on quantitative structure–activity relationships and an ensemble of neural networks (mtc-QSAR-EL) able to perform virtual screening for the identification of multi-strain inhibitors of *Mtb*. We also provide the physicochemical and fragment-based structural interpretation of the different molecular descriptors used to build the mtc-QSAR-EL model. Finally, we performed a virtual screening of a large and heterogeneous dataset of agency-regulated chemicals, identifying both already proven anti-TB agents and promising chemical structures with the potential to inhibit *Mtb*.

2. Results and Discussion

2.1. Performance of the Mtc-QSAR-EL Model and Applicability Domain

The best mtc-QSAR-EL model found by us has twelve $D[LQI]_{cj}$ descriptors, six of them based on hydrophobic factors, five containing steric information, and one focused on polar features of the molecules. Additionally, in terms of atom types, five $D[LQI]_{cj}$ descriptors were based on the effect of halogen atoms, three indicating the influence of heteroatoms (mainly N, O, S, and P), two characterizing the presence of methyl groups, and two accounting for the importance of aromatic carbons. The details regarding each $D[LQI]_{cj}$ descriptor appear in Table 1, while all the molecular descriptors of the chemicals and the corresponding biological data are reported in Supplementary Material 1.

The most appropriate mtc-QSAR-EL model developed here was an ensemble formed by three MLP networks whose parameters are represented in Table 2. These MLP networks have different numbers of neurons in the hidden layer and require diverse error functions and different numbers of epochs to be trained. Two of these MLP networks (first and third) have the same type of activation function (hyperbolic tangent), with the other one having a logistic function. In the output layer, the first and second MLP networks use a softmax function, while the third one uses a logistic function. The combination of these aspects resulted in a difference in performance among these MLP networks, particularly in the case of the sensitivities $[Sn(\%)]_{ta}$, $[Sn(\%)]_{bs}$, and $[Sn(\%)]_{ap}$, as well as the specificities $[Sp(\%)]_{ta}$, $[Sp(\%)]_{bs}$, and $[Sp(\%)]_{ap}$. These six local metrics were of paramount importance in assessing the classification performance of the mtc-QSAR-EL model in both trained and unseen data (training and test sets, respectively) when considering the dissimilar

experimental conditions c_j (see Section 3 for full explanation) under which the molecules were assayed.

Table 1. Descriptors of the type $D[LQI]c_j$ which entered in the mtc-QSAR-EL model.

Symbol	Definition
$D[LASSq3(HYD)G]ta$	Deviation of the atom-based stochastic quadratic index of order 3, weighted by the hydrophobicity of the halogens and their chemical environments, and depending on the chemical structure and the assay time.
$D[LASSq0(POL)G]ta$	Deviation of the atom-based stochastic quadratic index of order 0, weighted by the polarizability of the halogens, and depending on the chemical structure and the assay time.
$D[LASSq0(HYD)Y]ta$	Deviation of the atom-based stochastic quadratic index of order 0, weighted by the hydrophobicity of the heteroatoms, and depending on the chemical structure and the assay time.
$D[LASSq1(PSA)Y]ta$	Deviation of the atom-based stochastic quadratic index of order 1, weighted by the polar surface area of the heteroatoms and their adjacent atoms, and depending on the chemical structure and the assay time.
$D[LASSq0(HYD)M]bs$	Deviation of the atom-based stochastic quadratic index of order 0, weighted by the hydrophobicity of the methyl groups, and depending on the chemical structure and the <i>Mtb</i> strain against which the molecule was tested.
$D[LASSq0(AW)P]bs$	Deviation of the atom-based stochastic quadratic index of order 0, weighted by the atomic weight of the aromatic carbons, and depending on the chemical structure and the <i>Mtb</i> strain against which the molecule was tested.
$D[LASSq2(HYD)Y]bs$	Deviation of the atom-based stochastic quadratic index of order 2, weighted by the hydrophobicity of the heteroatoms and their chemical environments, and depending on the chemical structure and the <i>Mtb</i> strain against which the molecule was tested.
$D[LASSq2(HYD)G]ap$	Deviation of the atom-based stochastic quadratic index of order 2, weighted by the hydrophobicity of the halogens and their chemical environments, and depending on the chemical structure and information regarding the assay protocol.
$D[LASSq5(AW)G]ap$	Deviation of the atom-based stochastic quadratic index of order 5, weighted by the atomic weight of the halogens and their chemical environments, and depending on the chemical structure and information regarding the assay protocol.
$D[LASSq3(KU)G]ap$	Deviation of the atom-based stochastic quadratic index of order 3, weighted by Kupchick's vertex degrees of the halogens and their chemical environments, and depending on the chemical structure and information regarding the assay protocol.
$D[LASSq2(HYD)M]ap$	Deviation of the atom-based stochastic quadratic index of order 2, weighted by the hydrophobicity of the methyl groups and their chemical environments, and depending on the chemical structure and information regarding the assay protocol.
$D[LASSq2(AW)P]ap$	Deviation of the atom-based stochastic quadratic index of order 2, weighted by the atomic weight of the aromatic carbons and their chemical environments, and depending on the chemical structure and information regarding the assay protocol.

Table 2. Parameters characterizing the different MLP networks that were employed to build the mtc-QSAR-EL model.

Network Notation ^a	Training Algorithm ^b	Error Function ^c	Hidden Activation ^d	Output Activation
MLP 12-45-2	BFGS 116	Entropy	Tanh	Softmax
MLP 12-33-2	BFGS 138	Entropy	Logistic	Softmax
MLP 12-41-2	BFGS 104	SOS	Tanh	Logistic

^a The number of input nodes appears before the first hyphen, the number of neurons in the hidden layer appears between the two hyphens, and the number appearing after the second hyphen indicates the number of classes that were predicted by the mtc-QSAR-EL model. ^b BFGS—Broyden–Fletcher–Goldfarb–Shanno algorithm; the number accompanying the symbol BFGS is the number of epochs used to train the MLP networks. ^c SOS—sum of squares. ^d Tanh—hyperbolic tangent.

The mtc-QSAR-EL model exhibited accuracies [$Ac(\%)$] of 93.41% and 85.79% in the training and test sets, respectively. Furthermore, the global statistical indices depicted in Table 3 support the good general performance of the mtc-QSAR-EL model. For instance, the sensitivity $Sn(\%)$ and the specificity $Sp(\%)$ have values higher than 90% in the training set. These two global statistical indices reached values higher than 85% in the test set. Additionally, the MCC values are greater than 0.7, and, given their closeness to the ideal value of one, we can infer that there is very strong convergence between the observed [$ATB_i(cj)$] and predicted [$Pred_ATB_i(cj)$] values of the categorical variable of anti-TB activity against the different *Mtb* strains. The classification results for all the molecules of our dataset are reported in Supplementary Material 2. The file of the MLP networks can be obtained upon request to the authors.

Table 3. Statistical metrics indicating the global performance of the mtc-QSAR-EL model.

SYMBOLS ^a	Training Set	Test Set
N_{Active}	602	201
CC_{Active}	572	172
$Sn(\%)$	95.02%	85.57%
$N_{Inactive}$	582	186
$CC_{Inactive}$	534	160
$Sp(\%)$	91.75%	86.02%
MCC	0.869	0.716

^a N_{Active} —number of molecules labeled as active; $N_{Inactive}$ —number of molecules annotated as inactive; CC_{Active} —number of molecules rightly classified/predicted as active; $CC_{Inactive}$ —number of molecules correctly classified/predicted as inactive; $Sn(\%)$ —sensitivity (percentage of molecules properly classified as active); $Sp(\%)$ —specificity (percentage of molecules properly classified as inactive); MCC —Matthews' correlation coefficient.

At the local level, the mtc-QSAR-EL model also had a good performance. In the case of the training set, the local metrics [$Sn(\%)ta$, [$Sn(\%)bs$, [$Sn(\%)ap$, [$Sp(\%)ta$, [$Sp(\%)bs$, and [$Sp(\%)ap$] were in the range 70–100%. The only exception was the assay time, 4 d (four days), which exhibited [$Sp(\%)ap$] = 56.25%. In the case of the test set, a similar result was achieved since the same six statistical metrics were in the interval 71–100%, except for [$Sp(\%)ap$] = 50% (assay time of four days) and [$Sn(\%)ta$] = 64% (assay time of five days), as well as [$Sp(\%)bs$] and [$Sp(\%)ap$], with values of 61.54% for the strain *Mtb* (H37Rv_NRF) and the assay protocol “LORA method”, respectively. The previously mentioned global statistical indices and the local metrics discussed here confirm the internal quality and predictive power of the mtc-QSAR-EL model. All the values of [$Sn(\%)ta$], [$Sn(\%)bs$], [$Sn(\%)ap$], [$Sp(\%)ta$], [$Sp(\%)bs$], and [$Sp(\%)ap$] can be found in Supplementary Material 2.

Last, we would like to highlight that, from a physicochemical and structural point of view, the present mtc-QSAR-EL model classified anti-TB drugs belonging to a wide

spectrum of chemical families (Figures 1 and 2) such as polyketides, ethylenediamine derivative, aminoglycosides, nitroimidazopyrans, fluoroquinolones, and diarylquinolines.

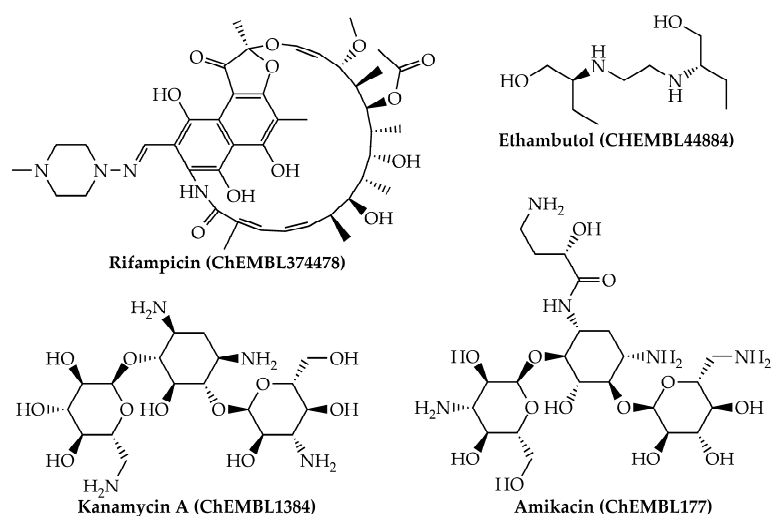


Figure 1. Anti-TB drugs (polyketides, ethylenediamine derivatives, and aminoglycosides) present in our dataset which were correctly classified by the mtc-QSAR-EL model.

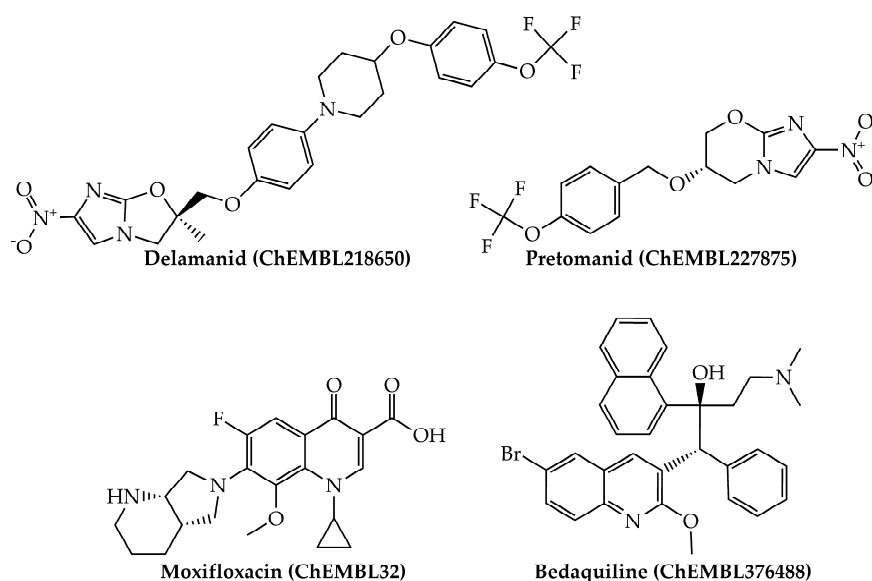


Figure 2. Other anti-TB drugs (nitroimidazopyrans, fluoroquinolones, and diarylquinolines) present in our dataset which were also correctly classified by the mtc-QSAR-EL model.

Such heterogeneity of chemical structures together with the different definitions of the $D[LQI]c_j$ descriptors, the size of the dataset, and the relatively high values of the global and local statistical indices points to the capability and appropriateness of the mtc-QSAR-EL model to predict the anti-TB activity of structurally dissimilar chemicals against the different *Mtb* strains.

Regarding the AD of the mtc-QSAR-EL model, as reported in seminal references (see Section 3.3), we calculated the so-called total score of applicability domain (TSAD), which is derived from the descriptors' space approach. Since our mtc-QSAR-EL model contained twelve $D[LQI]c_j$ descriptors, only those chemicals with TSAD = 12 were considered to be inside the AD (Supplementary Material 3). By carefully inspecting our dataset, we found that only 15 out of 1571 molecules/cases in the dataset were outside the AD of the mtc-QSAR-EL model, 14 of them with TSAD = 11 and one with TSAD = 10. However,

13 out of these 15 seemingly atypical chemicals were correctly classified by considering the consensus prediction approach performed by the mtc-QSAR-EL model. We decided to keep these chemicals in the dataset since the consensus predictions constituted the priority over the descriptors' space approach to define the AD.

2.2. Molecular Descriptors and Their Physicochemical and Structural Meanings

The sensitivity values *SV* of the *D[LQI]cj* descriptors are depicted in Figure 3, where the highest values indicate those *D[LQI]cj* descriptors with the greatest influence and discriminatory power in the mtc-QSAR-EL model. Simultaneously, the comparison between the class-based means for each *D[LQI]cj* descriptor represented in Table 4 shows us the type of variation that the value of a given *D[LQI]cj* descriptor should undergo to potentiate the anti-TB activity against the different *Mtb* strains.

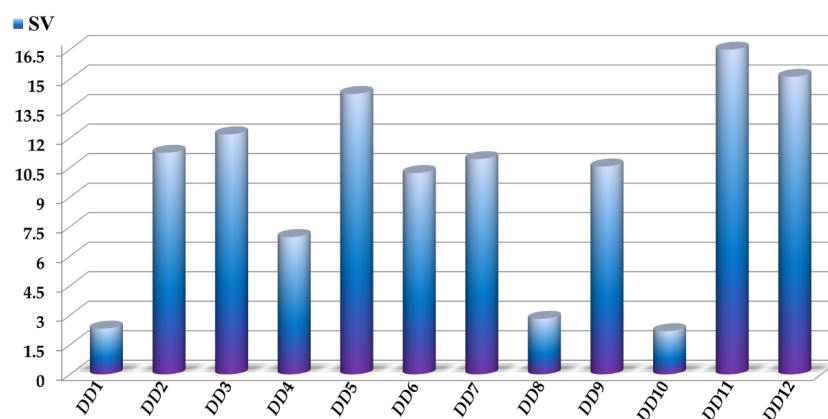


Figure 3. Molecular descriptors present in the mtc-QSAR-EL model and their relative influences assessed as sensitivity values (*SV*). For the sake of simplicity, we used the following abbreviations: DD1 = *D[LASSq3(HYD)G]ta*, DD2 = *D[LASSq0(POL)G]ta*, DD3 = *D[LASSq0(HYD)Y]ta*, DD4 = *D[LASSq1(PSA)Y]ta*, DD5 = *D[LASSq0(HYD)M]bs*, DD6 = *D[LASSq0(AW)P]bs*, DD7 = *D[LASSq2(HYD)Y]bs*, DD8 = *D[LASSq2(HYD)G]ap*, DD9 = *D[LASSq5(AW)G]ap*, DD10 = *D[LASSq3(KU)G]ap*, DD11 = *D[LASSq2(HYD)M]ap*, and DD12 = *D[LASSq2(AW)P]ap*.

Table 4. Tendencies of variation calculated for the different molecular descriptors in the mtc-QSAR-EL model according to the class-based mean approach.

Symbol	Class-Based Means		Tendency ^a
	Active	Inactive	
<i>D[LASSq3(HYD)G]ta</i>	-1.1976×10^{-4}	-2.0514×10^{-2}	Increase
<i>D[LASSq0(POL)G]ta</i>	5.0005×10^{-3}	6.3674×10^{-3}	Decrease
<i>D[LASSq0(HYD)Y]ta</i>	5.1000×10^{-3}	-4.4105×10^{-1}	Increase
<i>D[LASSq1(PSA)Y]ta</i>	1.2460×10^{-3}	-1.6800×10^{-1}	Increase
<i>D[LASSq0(HYD)M]bs</i>	1.4621×10^{-2}	-3.0065×10^{-1}	Increase
<i>D[LASSq0(AW)P]bs</i>	8.1935×10^{-3}	-2.2074×10^{-1}	Increase
<i>D[LASSq2(HYD)Y]bs</i>	-4.0897×10^{-3}	-2.1845×10^{-1}	Increase
<i>D[LASSq2(HYD)G]ap</i>	1.0793×10^{-2}	-2.4162×10^{-1}	Increase
<i>D[LASSq5(AW)G]ap</i>	1.2911×10^{-2}	-1.8708×10^{-1}	Increase
<i>D[LASSq3(KU)G]ap</i>	9.6900×10^{-3}	-1.1239×10^{-1}	Increase
<i>D[LASSq2(HYD)M]ap</i>	1.6609×10^{-2}	-2.8205×10^{-1}	Increase
<i>D[LASSq2(AW)P]ap</i>	1.1549×10^{-2}	-2.8297×10^{-1}	Increase

^a This denotes how the value of a *D[LQI]cj* descriptor should vary (increase or diminution) in order for a molecule to enhance its anti-TB activity and versatility (ability to inhibit more than one *Mtb* strain).

As mentioned before, in our mtc-QSAR-EL model, we have six $D[LQI]_{cj}$ descriptors characterizing information regarding atomic hydrophobic contributions. In this sense, we would like to highlight that such contributions are based on the hydrophobicity scale proposed by Ghose and co-workers [35]. According to this scale, aliphatic carbon atoms present negative hydrophobicity values, excluding the fragments of the form CHX_3 , CR_2X_2 , CRX_3 , and CX_4 ($X = O, N, S, P, Se,$ or halogen). Nitrogenated and oxygenated functional groups also have negative hydrophobic contributions except for the pyrrolic nitrogen (oxygen in furan) atom, $Ar-NH-Ar$ (and its oxygenated counterpart), with Ar being an aromatic (or heteroaromatic ring), and all the tertiary amines.

With that being said, the results presented in Table 4 indicate that the anti-TB activity through the inhibition of different *Mtb* strains can be enhanced by increasing the value of $D[LASSq3(HYD)G]_{ta}$, which describes the augmentation of the product of the hydrophobic contributions of any two atoms (with at least one of them being a halogen) that are placed at the topological distance (number of bonds between two atoms without considering bond multiplicity) of three. Examples of fragments with this characteristic are the 5-(halomethyl)pyrimidines, including those with substitutions in positions 4 and/or 6. Notice that $D[LASSq3(HYD)G]_{ta}$ is the eleventh most important $D[LQI]_{cj}$ descriptor in the mtc-QSAR-EL model. We also have $D[LASSq0(HYD)Y]_{ta}$ whose increase is directly proportional to the number of heteroatoms in a molecule. In this sense, the presence of fragments such as $Ar-NO_2$ ($Ar =$ aromatic or heteroaromatic ring), primary amines, amides, hydroxyl groups (alcohol), thiols, thioethers, functional groups containing sulfur (with sp^2 hybridization) attached to a carbon atom, phosphite, and $R_3-P = X$ ($R =$ any group link through carbon, $X = O$ or S) will considerably increase the value of $D[LASSq0(HYD)Y]_{ta}$ (the fourth most important $D[LQI]_{cj}$ descriptor), favoring the anti-TB activity. The inhibitory activity against the *Mtb* strains seems to be enhanced by the augmentation of the number of methyl groups in the molecules, and such a structural variation is characterized by $D[LASSq0(HYD)M]_{bs}$, which is the third most significant $D[LQI]_{cj}$ descriptor in the mtc-QSAR-EL model. We also have $D[LASSq2(HYD)Y]_{bs}$ (ranked the sixth most influential $D[LQI]_{cj}$ descriptor), which indicates the increase in the product of the hydrophobic contributions of any two atoms (with at least one of them being a heteroatom) that are placed at the topological distance of two. Substructures such as pyrimidin-2-amine, 2-alkylpyrimidines, and urea, as well as aliphatic chains (or alicyclic moieties) attached to hydroxyl, amino, amide, and sulfonamide groups, will favorably increase the value of $D[LASSq2(HYD)Y]_{bs}$, with the subsequent improvement in the anti-TB activity. The presence of halogens seems to be of great importance in the increase in the inhibitory activity against the *Mtb* strains, and as in the case of $D[LASSq3(HYD)G]_{ta}$ (commented above), $D[LASSq2(HYD)G]_{ap}$ also positively contributes through the increment in the product of the hydrophobic contributions of any two atoms (with at least one of them being a halogen) that are placed at the topological distance of two. Thus, 5-halopyrimidines and 4,6-disubstituted halobenzene fragments increase the value of $D[LASSq2(HYD)G]_{ap}$, which is the tenth most significant $D[LQI]_{cj}$ descriptor. Last, we have $D[LASSq2(HYD)M]_{ap}$ as the most important $D[LQI]_{cj}$ descriptor in the mtc-QSAR-EL model. In this sense, $D[LASSq2(HYD)M]_{ap}$ involves the increase in the hydrophobic contributions of any two atoms (with at least one of them being a carbon from a methyl group) that are placed at the topological distance of two. Moieties such as 2-methylpyrimidines, as well as aliphatic chains (or cycloalkane rings) containing methyl groups, amides, methoxy groups, and toluene fragments, will increase the value of $D[LASSq2(HYD)M]_{ap}$, and therefore the anti-TB activity against the diverse *Mtb* strains reported here.

Our mtc-QSAR-EL model also has five $D[LQI]_{cj}$ descriptors associated with steric factors. In this context, $D[LASSq0(POL)G]_{ta}$ describes the diminution of the polarizability of the halogens. The value of $D[LASSq0(POL)G]_{ta}$ (having the fifth highest influence) can be decreased either by diminishing the number of halogens or by the presence of fluorine atoms. Consequently, halogens are usually important for the activity profiles of the molecules, and the case of the anti-TB activity is not an exception. Thus, functional

groups such as trifluoromethyl (and, to a lesser degree, dichloromethyl and bromomethyl), 1,2-difluorobenzene, and 2-chlorobenzene will decrease the $D[LASSq0(POL)G]ta$ value enough to favor the inhibitory potency against any *Mtb* strain. Two of the five steric $D[LQI]cj$ descriptors characterize the presence of the aromatic carbons in the molecules. On one hand, the increase in the value of $D[LASSq0(AW)P]bs$ is directly proportional to the increase in the number of aromatic carbons (e.g., benzene rings and polycyclic substructures such as naphthalene), while $D[LASSq2(AW)P]ap$, in addition to benefiting from the presence of aromatic carbons, is also favored by the presence of relatively heavy atoms (Cl, Br, and S) placed at the topological distance of two with respect to an aromatic carbon. Therefore, fragments that also increase the value $D[LASSq2(AW)P]ap$ (and, to a lesser degree, $D[LASSq0(AW)P]bs$) are halobenzene and benzenethiol. The descriptors $D[LASSq0(AW)P]bs$ and $D[LASSq2(AW)P]ap$ rank eighth and second among the most important $D[LQI]cj$ descriptors, respectively. Two other steric $D[LQI]cj$ descriptors take into account the presence of halogens. One of them is $D[LASSq5(AW)G]ap$ (exhibiting the seventh highest influence), which characterizes the increase in the atomic weight of any two atoms (with at least one of them being a halogen) placed at the topological distance of five or less. Fragments of the type $ZCX3$ (Z = any atom, X = Cl or Br) and 1,4-dihalobenzene can favorably increase the value of $D[LASSq5(AW)G]ap$. The other steric $D[LQI]cj$ descriptor is $D[LASSq3(KU)G]ap$ and expresses the increment in the atomic accessibility (ability to interact with atoms from other molecules) of any two atoms (with at least one of them being a halogen) placed at the topological distance of three. In this case, 1,2-dihalobenzene substructures (Cl and Br favored over F) and moieties such as $ZCX3$, 5-halopyrimidines, and 4,6-disubstituted halobenzene are examples of fragments that increase the value of $D[LASSq3(KU)G]ap$. In the mtc-QSAR-EL model, $D[LASSq3(KU)G]ap$ is ranked twelfth among all the $D[LQI]cj$ descriptors.

Finally, we have $D[LASSq1(PSA)Y]ta$ (with the ninth highest significance) characterizing the increase in the polar surface area of any two adjacent heteroatoms, and, therefore, only pyridazine and 1,2,3-triazine rings, as well as the sulfonamide and phosphorus-containing functional groups (phosphorus forming bonds with only oxygen and/or sulfur), will considerably augment the value of $D[LASSq1(PSA)Y]ta$.

2.3. Computational Drug Repurposing of Agency-Regulated Chemicals as Anti-TB Agents

We performed the virtual screening of a dataset formed by 8898 agency-regulated chemicals (Supplementary Material 4), which included (but was not limited to) investigational and FDA-approved drugs. These were predicted by considering the 24 experimental conditions cj reported in the dataset used to build the mtc-QSAR-EL model, yielding 213,552 predictions (Supplementary Materials 5 and 6). In terms of the reliability of the predictions using the AD of the mtc-QSAR-EL model according to the descriptors' space approach, we generated the TSAD values for each of these 8898 chemicals (Supplementary Material 7). Then, the metrics FA(%) and S(TSAD) were calculated (see Section 3.5 for full explanation). In any case, we would like to highlight that FA(%) describes the anti-TB potential of a molecule because it expresses the frequency in which a chemical is predicted as active against *Mtb*. A high FA(%) value (maximum value is 100%) for a chemical means that it has a higher probability of having anti-TB activity by inhibiting multiple *Mtb* strains with $MIC_{90} < 7622.22$ nM, which is the highest of the MIC_{90} cutoffs reported in this work (see Section 3.1). At the same time, a high S(TSAD) value (the ideal value = 12×24 experimental conditions $cj = 288$, with 12 being the maximum TSAD value) indicates the general tendency of a chemical to be inside the AD according to the descriptors' space approach, which, together with the consensus predictions performed by the mtc-QSAR-EL model, helps to assess the degree of reliability of such predictions.

The combined use of FA(%) and S(TSAD) (Supplementary Material 8) allowed us to rank the 8898 agency-regulated chemicals, and those with $FA(\%) > 80\%$ and $S(TSAD) \geq 276$ were the top ranked, giving priority to those exhibiting the highest FA(%) values. Notice that there is no "rule of thumb" in terms of the criteria used to prioritize chemicals. Therefore, in the present study, we employed these rigorous cutoff values for FA(%) and

S(TSAD) to achieve a virtual screening hit rate of 0.49% (44 out of 8898 chemicals) which is higher than the greatest hit rate value of 0.14% for high-throughput screening but lower than smallest hit rate value for virtual screening campaigns (1%) [36].

By using the aforementioned metrics for compound prioritization, the mtc-QSAR-EL model identified three chemicals with experimentally proven anti-TB activity (Figure 4): macozinone (PBTZ-169), BTZ-043, and niclosamide. Notice that macozinone is a remarkably potent piperazine-containing benzothiazinone, being able to inhibit multiple *Mtb* strains at $MIC_{99} \leq 1$ nM [37]. In the case of BTZ-043, although structurally related to macozinone, it lacks the 2-(4-(cyclohexylmethyl)piperazin-1-yl) moiety, which decreases its activity. Still, BTZ-043 is a nanomolar inhibitor of several *Mtb* strains [37,38]. On the other hand, niclosamide offers very attractive opportunities for anti-TB therapies because, in addition to being a recognized antihelminthic drug, it also has a wide-spectrum antimicrobial profile, which includes nanomolar to micromolar potency against diverse viruses (including coronaviruses) [39], as well as anti-TB activity in the low micromolar range [40,41].

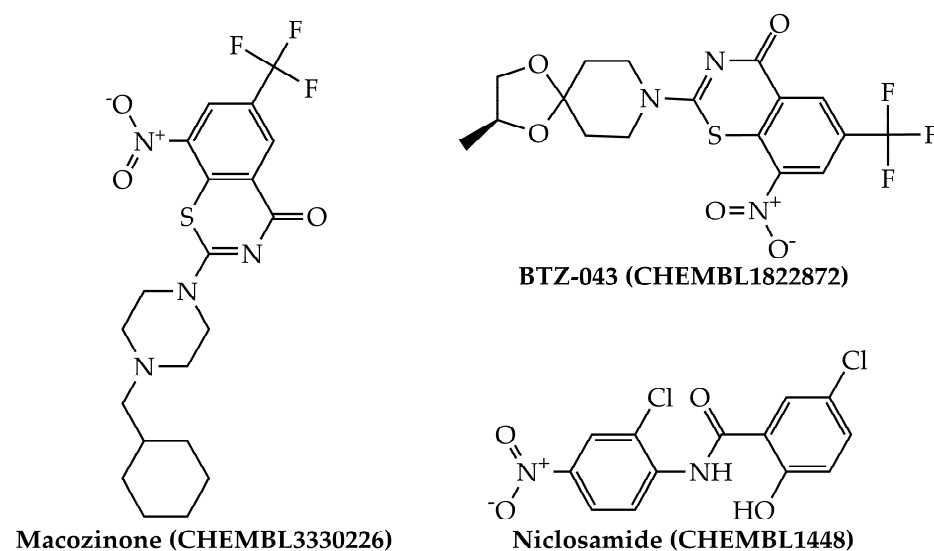


Figure 4. Structures of the agency-regulated chemicals with a proven anti-TB activity identified by the mtc-QSAR-EL model.

We should recall that the cutoff values of the metrics FA(%) and S(TSAD) used by us are very rigorous. If, for instance, we slightly relax these two metrics (e.g., FA(%) > 60% and S(TSAD) \geq 270), other agency-regulated chemicals with experimentally proven anti-TB activity pop up. These are the cases of biapenem (FA(%) = 66.67% and S(TSAD) = 288) and TBA-354 (FA(%) = 91.67% and S(TSAD) = 270), whose anti-TB profile has been demonstrated in vitro (low micromolar range) and in vivo [42,43]. Notice that further relaxing FA(%) and/or S(TSAD) will allow the mtc-QSAR-EL model to identify more anti-TB agents, but a larger number of false positives may also be predicted. In the end, it is up to the analyst to select the adequate values of the metrics FA(%) and S(TSAD), which will depend on the number of drugs available for testing, with this being a key element when planning the expenditure of financial resources for experimental validation of virtually selected chemicals. In any case, we advise the use of FA(%) > 29% and S(TSAD) \geq 250 since with these cutoffs, most of the known FDA-approved and investigational anti-TB drugs (not included in the dataset used to build our mtc-QSAR-EL model) were identified in the virtual screening analysis. We would like to highlight that the FA(%) value suggested by us is in the range reported for the hit rate in the prospective virtual screening [36,44].

Returning to the top 44 molecules predicted by the mtc-QSAR-EL model from the 8898 agency-regulated chemicals, we ran an experiment. To obtain a deeper insight regarding the new molecular patterns with promising anti-TB potential, we used the webserver mycoCSM [45], which is an online tool with the capability to predict the antimycobacterial

profile of any molecule, including the anti-TB activity. The results of the top 44 agency-regulated chemicals identified as potential anti-TB agents (multi-strain inhibitors of *Mtb*) by our mtc-QSAR-EL model together with the predictions derived from mycoCSM appear in Table 5. It can be seen that the mtc-QSAR-EL model and the webserver mycoCSM converge in 10 out of 44 chemicals (22.73%). In our opinion and experience, such a convergence is very good taking into account that the mtc-QSAR-EL model and the webserver mycoCSM employed dissimilar approaches to characterize the molecular structure, different machine learning algorithms, and distinct ways to consider experimental information when building the computational models. In the end, given all the experimental and theoretical evidence, we can conclude that our mtc-QSAR-EL model can be efficiently used alone or in combination with other in silico tools for virtual screening of anti-TB molecules, which may inhibit several *Mtb* strains.

Table 5. Top-ranked regulated chemicals identified by mtc-QSAR-EL as potential multi-strain inhibitors against *Mtb* (MIC₉₀ < 7622.22 nM) and predicted by the webserver mycoCSM.

ID ^a	Name	FA(%) ^b	S(TSAD) ^c	log ₁₀ (MIC) ^d	MIC (nM) ^e
CHEMBL78535	Ancriviroc	100.00	288	−4.624	23768.40
CHEMBL1450565	Aklomide	100.00	279	−4.128	74473.20
CHEMBL2104712	Nisterime acetate	100.00	279	−4.673	21232.44
CHEMBL2106822	Nizofenone	95.83	288	−4.332	46558.61
CHEMBL2104159	Bromoxanide	100.00	276	−4.838	14521.12
CHEMBL1199080	Bretylum	95.83	279	−3.91	123026.88
CHEMBL3330226	Macoazinone	95.83	279	−7.558	27.67
CHEMBL1909324	Pinaverium	95.83	279	−4.668	21478.30
CHEMBL292702	Maitansine	95.83	279	−4.776	16749.43
CHEMBL2111120	Nitralamine	95.83	276	−4.02	95499.26
CHEMBL289832	Licostinel	95.83	276	−3.826	149279.44
CHEMBL1448	Niclosamide	95.83	276	−5.249	5636.38
CHEMBL2104616	Clonitazene	91.67	288	−4.866	13614.45
CHEMBL1269025	Edoxaban	91.67	288	−5.731	1857.80
CHEMBL2106056	Cronidipine	91.67	288	−4.771	16943.38
CHEMBL1241348	Faldaprevir	91.67	288	−5.275	5308.84
CHEMBL2013174	Vedroprevir	91.67	288	−5.446	3580.96
CHEMBL2104730	Nitroxinil	91.67	279	−3.881	131522.48
CHEMBL493636	Sulfanitran	91.67	279	−4.678	20989.40
CHEMBL9484	Clofilium	91.67	279	−4.042	90782.05
CHEMBL1822872	BTZ-043	91.67	279	−7.01	97.72
CHEMBL1178725	Nolpitantium	91.67	279	−4.876	13304.54
CHEMBL2104390	Ilatreotide	91.67	279	−5.123	7533.56
CHEMBL56367	Nimesulide	87.50	288	−4.907	12387.97
CHEMBL2107448	Loprazolam	87.50	288	−4.945	11350.11
CHEMBL3670800	ALK-4290	87.50	288	−5.298	5035.01
CHEMBL1181731	Teglarinad	87.50	288	−5.154	7014.55
CHEMBL1170047	Iniparib	87.50	279	−4.117	76383.58
CHEMBL491	Hydroxyflutamide	87.50	279	−4.221	60117.37
CHEMBL452	Clonazepam	87.50	279	−4.223	59841.16
CHEMBL1274	Nilutamide	87.50	279	−4.559	27605.78
CHEMBL2110930	Fubrogonium	87.50	279	−4.125	74989.42
CHEMBL2110825	Dodeclonium	87.50	279	−4.19	64565.42
CHEMBL397647	JNJ-17166864	87.50	279	−4.53	29512.09
CHEMBL2105721	Nivocasan	87.50	279	−4.644	22698.65
CHEMBL2107326	Dasantafil	87.50	279	−5.015	9660.51
CHEMBL1908326	Meclonazepam	83.33	288	−4.281	52360.04
CHEMBL1823817	CE-224535	83.33	288	−4.736	18365.38
CHEMBL1276663	Cefozopran	83.33	288	−5.143	7194.49
CHEMBL2110800	Ciclonium	83.33	279	−4.48	33113.11
CHEMBL1337	Nitisinone	83.33	279	−4.569	26977.39
CHEMBL512306	[18F]D	83.33	279	−4.532	29376.50
CHEMBL435191	Edotecarin	83.33	279	−4.521	30130.06
CHEMBL2074922	Efonidipine	83.33	279	−4.894	12764.39

^a Highlighted chemicals are those predicted by the mtc-QSAR-EL model and the webserver mycoCSM to have MIC₉₀ < 7622.22 nM.

^b FA(%)—percentage of times that a chemical was predicted as active (anti-TB agent) by the mtc-QSAR-EL model. ^c S(TSAD)—sum of all the TSAD values for a given chemical by considering all the experimental conditions reported in this study. ^d log₁₀(MIC)—predicted value of anti-TB activity estimated by the webserver mycoCSM. ^e Minimum inhibitory concentration calculated from the predicted log₁₀(MIC).

3. Materials and Methods

3.1. Dataset and Computation of the Molecular Descriptors

All the chemical and biological data associated with the anti-TB activity were retrieved from the public database known as ChEMBL [46–48]. The dataset used in the present study was formed by 1237 molecules belonging to different chemical families. These molecules were experimentally tested for their inhibitory activity against *Mtb* and where the MIC₉₀ (minimum inhibitory concentration that prevents the visible growth in 90% of the *Mtb* isolates) was measured. More specifically, each molecule was assayed by considering at least 1 out of 7 assay times (*ta*), against at least 1 out of 8 *Mtb* strains (*bs*), and involving at least 1 out of 4 assay protocols (*ap*). Notice that each combination of the elements *ta*, *bs*, and *ap* represents a unique experimental condition *c_j*, which can be expressed as *c_j*(*ta*, *bs*, *ap*). If a molecule was found to be assayed more than one time by considering the same experimental condition *c_j*, the duplicate data were deleted, and we kept the lowest MIC₉₀ value for that molecule. If two stereoisomers were tested under the same experimental condition *c_j*, we kept only the stereoisomer with the lowest MIC₉₀ value. In any case, in our dataset, most of the molecules were tested by considering only one *c_j*, and, therefore, after removing entries with lacking SMILES codes, values, units, duplicates, and unclear information regarding the assay time or the test protocol, the dataset ended up having 1571 cases. Each case/molecule in the dataset was annotated as active ($ATB_i(c_j) = 1$) or inactive ($ATB_i(c_j) = -1$), where $ATB_i(c_j)$ was a dichotomous variable indicating the anti-TB activity of the *i*th case/molecule under the experimental condition *c_j*. Assignments of active and inactive cases were realized by considering the different MIC₉₀ cutoff values depending on the assay time (Table 6).

Table 6. Experimental conditions *c_j* (combinations of the elements *ta*, *bs*, and *ap*) reported in the present work.

MIC ₉₀ Cutoff Value (nM) ^a	<i>ta</i> ^b	<i>bs</i> ^{c,d}	<i>ap</i> ^e
<1146.75	10d	<i>Mtb</i> (H37Rv_NRF)	LORA method
	10d	<i>Mtb</i> (H37Rv)	Spectrophotometric assay (OD580-OD600)
≤1500	14d	<i>Mtb</i> (H37Rv)	Broth dilution method
<7622.22	3d	<i>Mtb</i> (H37Rv_NRF)	Spectrophotometric assay (OD580-OD600)
	3d	<i>Mtb</i> (MC2 6220_NRF)	Spectrophotometric assay (OD580-OD600)
	3d	<i>Mtb</i> (MC2 6220_RF)	Spectrophotometric assay (OD580-OD600)
<5829.77	4d	<i>Mtb</i> (H37Rv)	AlamarBlue/Resazurin/MABA method
≤5300	5d	<i>Mtb</i> (H37Rv)	AlamarBlue/Resazurin/MABA method
	5d	<i>Mtb</i> (H37Rv_ATCC 25618)	Spectrophotometric assay (OD580-OD600)
	5d	<i>Mtb</i> (H37Rv)	Spectrophotometric assay (OD580-OD600)
	5d	<i>Mtb</i> (INH-R)	AlamarBlue/Resazurin/MABA method
	5d	<i>Mtb</i> (H37Rv_ATCC 27294)	Broth dilution method
≤5000	6d	<i>Mtb</i> (H37Rv_ATCC 25618)	AlamarBlue/Resazurin/MABA method
	6d	<i>Mtb</i> (H37Rv)	AlamarBlue/Resazurin/MABA method
	6d	<i>Mtb</i> (MC2 6220_NRF)	Spectrophotometric assay (OD580-OD600)
<4940	7d	<i>Mtb</i> (H37Rv)	AlamarBlue/Resazurin/MABA method
	7d	<i>Mtb</i> (H37Rv_ATCC 27294)	AlamarBlue/Resazurin/MABA method
	7d	<i>Mtb</i> (INH-R)	AlamarBlue/Resazurin/MABA method
	7d	<i>Mtb</i> (H37Rv_ATCC 27294)	Broth dilution method
	7d	<i>Mtb</i> (H37Rv)	Broth dilution method
	7d	<i>Mtb</i> (RMP-R)	AlamarBlue/Resazurin/MABA method
	7d	<i>Mtb</i> (H37Rv)	Spectrophotometric assay (OD580-OD600)
	7d	<i>Mtb</i> (MC2 6220_RF)	Spectrophotometric assay (OD580-OD600)
	7d	<i>Mtb</i> (MC2 6220_NRF)	Spectrophotometric assay (OD580-OD600)

^a Activity value from which a molecule was considered active. ^b *ta*—assay time. ^c *bs*—bacterial strain belonging to *Mtb*. ^d The following abbreviations were used: RF (non-replicative form), RF (replicative form), INH-R (isoniazid-resistant strain), and RMP-R (rifampicin-resistant strain). ^e *ap*—information regarding the assay protocol.

Several reasons justified the selection of the MIC₉₀ cutoffs in Table 6. First, by inspecting our dataset formed by the 1571 cases, one can see that there are FDA-approved anti-TB drugs (isoniazid, rifampicin, ethambutol, etc.) with MIC₉₀ values that fall either

above or below the cutoffs, which means that if a query chemical is predicted as active (or inactive) by the mtc-QSAR-EL model, it will be possible to assess the differences between the inhibitory potential of that query molecule and the actual activity of the current anti-TB drugs as a consequence of their differences in their chemical structures. Second, the annotations of the molecules as active or inactive employed a classification approach instead of a regression one. Notice that, in contrast to their regression counterparts, classification approaches do not need to predict the exact values of a biological property in a dataset, and, therefore, they do not need to deal so much with the potentially great uncertainty of the data. Third, the chosen MIC₉₀ cutoffs prevented (as much as it was possible) the imbalance between the number of molecules labeled as active and the number of molecules considered inactive. Fourth, the MIC₉₀ cutoffs are in the range (some of them are lower and, therefore, more rigorous) of the cutoffs selected in high-throughput screening campaigns to prioritize chemicals with potent anti-TB activity [49]. Fifth, from a phenomenological point of view, even though several anti-TB drugs appear in our dataset, the purpose of our mtc-QSAR-EL was to perform virtual screening of large and heterogenous external datasets to identify novel molecular patterns different from those present in the current anti-TB drugs.

The SMILES codes of the 1571 cases were stored in a file of the type *.smi. Following this, we used the software OpenBabel v2.4.0 [50] to convert this file to *.sdf, obtaining the connectivity table for each chemical present in our dataset; no additional standardization actions were applied. Then, using the sdf file as input, we employed the software QuBiLS-MAS v1.0 to compute the molecular descriptors known as local atom-based stochastic quadratic indices *LQI* [51,52]. These are topological descriptors with successful applications in medicinal chemistry and drug discovery [53,54]. We calculated the *LQI* descriptors of order *k* (with *k* from 0 to 5) by using predefined parameters such as algebraic form (quadratic), constraints (atom-based), matrix type (stochastic), cutoff (keep all), groups (local—referring to specific atom types such as aliphatic and aromatic carbons, methyl groups, halogens, and heteroatoms), and aggregation operator (Manhattan distance). These *LQI* descriptors were weighted by atomic properties such as hydrophobicity (*HYD*), atomic weight (*AW*), polarizability (*POL*), polar surface area (*PSA*), and Kupchick vertex degree (*KU*).

Notice that the *LQI* descriptors are not able to discriminate the effect of the chemical structure of a molecule on the anti-TB activity when the experimental condition *cj* changes, e.g., use of different assay times (*ta*), *Mtb* strains (*bs*), and/or test protocols (*ap*). This means that the *LQI* descriptor for a molecule will have the same value regardless of the experimental condition *cj* used to assess the anti-TB activity of that molecule. To solve this inconvenience, we applied the adaptation of the Box–Jenkins approach, which is a distinctive characteristic of all the PTML models [18–34,55]:

$$avg[LQI]_{cj} = \frac{1}{n(cj)} \times \sum_{i=1}^{n(cj)} LQI_i \quad (1)$$

In Equation (1), the term *avg[LQI]cj* is the average of the *LQI_i* values for all the molecules in the training set, which were annotated as active by considering the same experimental condition *cj*. Consequently, *n(cj)* refers to the number of molecules labeled as active (also in the training set) that were tested under the same *cj*. Notice that, as mentioned before, the experimental condition *cj* depends on the elements *ta*, *bs*, and *ap*, and, therefore, Equation (1) was applied to each of the elements of *cj* separately. For instance, if *cj* = *ta*, then *avg[LQI]cj* = *avg[LQI]ta* and *n(cj)* = *n(ta)*, meaning that, in this case, Equation (1) was employed for calculations depending only on the assay times. The same line of thinking was applied to the elements *bs* and *ap*. In the second step of the Box–Jenkins approach,

$$D[LQI]_{cj} = \left[\frac{LQI - avgLQI(cj)}{SDev(LQI)} \right] \cdot \sqrt{p(cj)} \quad (2)$$

In Equation (2), $D[LQI]_{cj}$ is a multi-target descriptor that considers both the chemical structure of a molecule and a specific element of the experimental condition cj . Therefore, as in the case of Equation (1), Equation (2) was applied to each element of cj separately. The descriptor $D[LQI]_{cj}$ characterizes how much any molecule physicochemically and structurally deviates from a group of molecules annotated as active, having been tested by considering the same element of cj . Additionally, $SDev[LQI]$ is the standard deviation calculated from all the values of each LQI descriptor for the molecules/cases present only in the training set. Last, the term $p(cj)$ is the a priori probability calculated as the quotient of the number of molecules/cases in the training set assayed by involving a given element of cj and the total number of compounds present in the training set.

3.2. Building the Mtc-QSAR-EL Model

When generating the mtc-QSAR-EL model, we followed a series of steps. First, we divided our dataset (containing 1571 cases) into training and test sets. In this sense, for each assay time, we sorted the molecules in ascending order of their MIC_{90} values. After, within each assay time, the first three molecules/cases were annotated as members of the training set, while the fourth molecule/case was labeled to belong to the test set. Such a ratio of 3:1 was repeated in the whole dataset. Thus, the training set was formed by 1184 molecule/cases (75.37% of the dataset), 602 active and 582 inactive; the training set was employed to find the best mtc-QSAR-EL model. The test set (the remaining 24.63%) contained 387 molecules/cases, 201 considered active and 186 defined as inactive; the test set was used to estimate the predictive power of the mtc-QSAR-EL model.

We employed the software IMMAN v1.0 [56], which allowed us to prioritize those $D[LQI]_{cj}$ descriptors that were likely to exhibit the highest discriminatory power. In doing so, we used two information theory metrics: the differential Shannon entropy [57] and the information gain ratio [58]. We ranked the $D[LQI]_{cj}$ descriptors according to their decreasing values of the geometric mean between the two aforementioned metrics. While selecting the $D[LQI]_{cj}$ descriptors with the highest discriminatory power (high geometric mean values), we estimated the redundancy among them by computing the pairwise Pearson correlation coefficient (PCC) [59]. Only the $D[LQI]_{cj}$ descriptors with pairwise correlation values in the interval $-0.7 < PCC < 0.7$ were chosen to construct the mtc-QSAR-EL model.

We should recall that the mtc-QSAR-EL model is an ensemble of artificial neural networks (ANNs), and, thus, to form such an ensemble, we searched for the best ANNs, all of them being based on the multi-layer perceptron (MLP) architecture. When selecting the most appropriate MLP networks, we examined, in a first step, global statistical metrics such as accuracy [$Ac(\%)$], Matthews' correlation coefficient (MCC) [60], sensitivity [$Sn(\%)$], and specificity [$Sp(\%)$]. However, ultimately, the best mtc-QSAR-EL model was chosen as an ensemble of those MLP networks that exhibited the highest values of the sensitivities [$Sn(\%)_{ta}$], [$Sn(\%)_{bs}$], and [$Sn(\%)_{ap}$], as well as the specificities [$Sp(\%)_{ta}$], [$Sp(\%)_{bs}$], and [$Sp(\%)_{ap}$]. Notice that these six local statistical indices depended on specific elements of the experimental condition cj . The ANN package of the software STATISTICA v13.5.0.17 [61] was employed to generate the MLP networks and subsequently build the mtc-QSAR-EL model.

3.3. Applicability Domain

When defining the applicability domain (AD) of the mtc-QSAR-EL model, we used two approaches. One of them was the consensus predictions since our mtc-QSAR-EL model was an ensemble of different MLP networks [62,63]. This means that for each molecule present in our dataset, the predicted probabilities given by each of the MLP networks were averaged, resulting in the probability of the ensemble (mtc-QSAR-EL) model for that molecule. As the second method to estimate the AD, we used a modification of the descriptors' space approach reported by Speck-Planche [64]. In doing so, for each molecule present in the dataset containing the 1571 cases/molecules, we generated the local scores

(LS) of the applicability domain for each $D[LQI]_{cj}$ descriptor in the following manner. If for a specific $D[LQI]_{cj}$ descriptor, the $D[LQI]_{cj}$ value of a query molecule fell between the maximum and minimum $D[LQI]_{cj}$ values, the LS took the value of one; otherwise, the LS was equal to zero. This operation was repeated for each $D[LQI]_{cj}$ descriptor that entered in the mtc-QSAR-EL model. Then, the metric known as the total score of applicability domain (TSAD) was calculated for the aforementioned query molecule as the sum of the LS values. In the end, in magnitude, the maximum value of TSAD was equal to the number of $D[LQI]_{cj}$ descriptors present in the mtc-QSAR-EL model.

3.4. Interpretation of the Molecular Descriptors in the Mtc-QSAR-EL Model

Due to the non-linear nature of the mtc-QSAR-EL model, we provided the physico-chemical and structural interpretation of the $D[LQI]_{cj}$ descriptors by strictly following the guidelines recently reported by Speck-Planche and co-workers [21,64]. In this sense, we used the sensitivity values SV calculated by the ANN package of the software STATISTICA v13.5.0.17 to rank the different $D[LQI]_{cj}$ descriptors in terms of their importance in the mtc-QSAR-EL model while calculating the class-based means to help us gain insights regarding how the values of the $D[LQI]_{cj}$ descriptors should vary (increase or decrease) to enhance both the anti-TB activity and the versatility of a molecule to inhibit more than one *Mtb* strain.

3.5. Virtual Screening of Agency-Regulated Chemicals

We virtually screened a large and heterogeneous dataset formed by 8898 agency-regulated chemicals. None of these molecules were present in the dataset used to build the mtc-QSAR-EL model (1571 chemicals/cases). We predicted each of the 8898 agency-regulated chemicals under the 24 experimental conditions c_j reported in Table 6. Then, we generated the metric $FA(\%)$, which expressed the percentage of experimental conditions c_j in which each of these 8898 agency-regulated chemicals was predicted as active (anti-TB agent) by the mtc-QSAR-EL model (Supplementary Material 8). We also defined a second metric symbolized as $S(TSAD)$; this was calculated as the sum of all the TSAD values reported for a single agency-regulated chemical by considering the 24 experimental conditions c_j . The meaning of $S(TSAD)$ was that the higher its value, the more reliable the predictions associated with that agency-regulated chemical in terms of whether it fell within the AD of the mtc-QSAR-EL model by considering the 24 experimental conditions c_j (Supplementary Material 8). By using $FA(\%)$ and $S(TSAD)$, we could rank the agency-regulated chemicals according to their predicted anti-TB activity and the reliability of the predictions, respectively.

4. Conclusions

The search for novel and more efficient anti-TB therapies can be greatly accelerated by employing *in silico* models as tools in the context of computational drug repurposing, where the prioritized hits could then be experimentally validated. Current computer-aided drug discovery approaches should, however, focus on including more information regarding the experimental conditions under which the molecules are assayed. By doing so, computational models will be able to make more accurate predictions while also providing a deeper phenomenological understanding of the physicochemical properties and structural requirements associated with the potentiation of the anti-TB activity and versatility to inhibit different *Mtb* strains. The mtc-QSAR-EL model created in this work constitutes an advance in early drug discovery against TB, demonstrating that it is possible to identify anti-TB agents and/or multi-strain inhibitors of *Mtb* via virtual screening while proposing new molecular patterns that could be promising as starting points for anti-TB drug development. The present report confirms the promising applications of the PTML modeling methodology, which can be extended to diverse research areas devoted to finding treatments for unmet needs.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/antibiotics10081005/s1>. Supplementary Material 1: Chemical and biological data including the *LQI* descriptors and the averages. Supplementary Material 2: *D[LQI]cj* descriptors, classification results, and local metrics. Supplementary Material 3: Applicability domain of the mtc-QSAR-EL model. Supplementary Material 4: Chemical data of dataset containing the 8898 agency-regulated chemicals and their *LQI* descriptors. Supplementary Material 5: *D[LQI]cj* descriptors calculated for the 8898 agency-regulated chemicals. Supplementary Material 6: Prediction results for the 8898 agency-regulated chemicals under multiple experimental conditions. Supplementary Material 7: Applicability domain of the predictions involving the 8898 agency-regulated chemicals. Supplementary Material 8: Calculation of the metrics FA(%) and S(TSAD).

Author Contributions: Conceptualization, A.S.-P.; methodology, A.S.-P.; software, A.S.-P. and V.V.K.; validation, A.S.-P.; formal analysis, A.S.-P.; investigation, A.S.-P., V.V.K. and M.T.S.; resources, A.S.-P., V.V.K. and M.T.S.; data curation, V.V.K.; writing—original draft preparation, A.S.-P., V.V.K. and M.T.S.; writing—review and editing, A.S.-P.; visualization, A.S.-P., V.V.K. and M.T.S.; supervision, A.S.-P.; project administration, A.S.-P., V.V.K. and M.T.S.; funding acquisition, M.T.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Brazilian National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico—CNPq) under the grant numbers 309648/2019-0 and 431254/2018-4. The APC was fully waived.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the chemical and biological (raw) data were retrieved from the public repository known as ChEMBL (<https://www.ebi.ac.uk/chembl/>, accessed on 23 June 2021).

Acknowledgments: The authors would like to acknowledge all the experimental scientists working in the field of antituberculosis research, whose data made the creation of our mtc-QSAR-EL model possible.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. WHO Global Tuberculosis Report. Available online: https://www.who.int/tb/publications/global_report/tb19_Exec_Sum_12_Nov2019.pdf?ua=1 (accessed on 20 July 2021).
2. Louw, G.E.; Warren, R.M.; Gey van Pittius, N.C.; McEvoy, C.R.; Van Helden, P.D.; Victor, T.C. A balancing act: Efflux/influx in mycobacterial drug resistance. *Antimicrob. Agents Chemother.* **2009**, *53*, 3181–3189. [CrossRef]
3. Grzelak, E.M.; Choules, M.P.; Gao, W.; Cai, G.; Wan, B.; Wang, Y.; McAlpine, J.B.; Cheng, J.; Jin, Y.; Lee, H.; et al. Strategies in anti-Mycobacterium tuberculosis drug discovery based on phenotypic screening. *J. Antibiot.* **2019**, *72*, 719–728. [CrossRef] [PubMed]
4. Hoagland, D.T.; Liu, J.; Lee, R.B.; Lee, R.E. New agents for the treatment of drug-resistant Mycobacterium tuberculosis. *Adv. Drug Deliv. Rev.* **2016**, *102*, 55–72. [CrossRef] [PubMed]
5. DiMasi, J.A.; Grabowski, H.G.; Hansen, R.W. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J. Health Econ.* **2016**, *47*, 20–33. [CrossRef]
6. Pushpakom, S.; Iorio, F.; Eyers, P.A.; Escott, K.J.; Hopper, S.; Wells, A.; Doig, A.; Guilliams, T.; Latimer, J.; McNamee, C.; et al. Drug repurposing: Progress, challenges and recommendations. *Nat. Rev. Drug Discov.* **2019**, *18*, 41–58. [CrossRef]
7. Speck-Planche, A.; Cordeiro, M.N.D.S. Simultaneous modeling of antimycobacterial activities and ADMET profiles: A chemoinformatic approach to medicinal chemistry. *Curr. Top. Med. Chem.* **2013**, *13*, 1656–1665. [CrossRef]
8. Speck-Planche, A.; Kleandrova, V.V.; Cordeiro, M.N.D.S. New insights toward the discovery of antibacterial agents: Multi-tasking QSBER model for the simultaneous prediction of anti-tuberculosis activity and toxicological profiles of drugs. *Eur. J. Pharm. Sci.* **2013**, *48*, 812–818. [CrossRef]
9. Speck-Planche, A.; Kleandrova, V.V.; Luan, F.; Cordeiro, M.N.D.S. In silico discovery and virtual screening of multi-target inhibitors for proteins in Mycobacterium tuberculosis. *Comb. Chem. High Throughput Screen.* **2012**, *15*, 666–673. [CrossRef]
10. Speck-Planche, A.; Scotti, M.T.; García-López, A.; Emerenciano, V.P.; Molina-Pérez, E.; Uriarte, E. Design of novel antituberculosis compounds using graph-theoretical and substructural approaches. *Mol. Divers.* **2009**, *13*, 445–458. [CrossRef] [PubMed]
11. Prathipati, P.; Ma, N.L.; Keller, T.H. Global Bayesian models for the prioritization of antitubercular agents. *J. Chem. Inf. Model.* **2008**, *48*, 2362–2370. [CrossRef] [PubMed]
12. Sivakumar, P.M.; Geetha Babu, S.K.; Mukesh, D. QSAR studies on chalcones and flavonoids as anti-tuberculosis agents using genetic function approximation (GFA) method. *Chem. Pharm. Bull.* **2007**, *55*, 44–49. [CrossRef]

13. Ragno, R.; Marshall, G.R.; Di Santo, R.; Costi, R.; Massa, S.; Rompei, R.; Artico, M. Antimycobacterial pyrroles: Synthesis, anti-MycoBacterium tuberculosis activity and QSAR studies. *Bioorg. Med. Chem.* **2000**, *8*, 1423–1432. [[CrossRef](#)]
14. Nesměrák, K.; Toropov, A.A.; Toropova, A.P.; Ertan-Bolelli, T.; Yildiz, I. QSAR of antimycobacterial activity of benzoxazoles by optimal SMILES-based descriptors. *Med. Chem. Res.* **2017**, *26*, 3203–3208. [[CrossRef](#)]
15. Ortega-Tenezaca, B.; Quevedo-Tumaili, V.; Bediaga, H.; Collados, J.; Arrasate, S.; Madariaga, G.; Munteanu, C.R.; Cordeiro, M.; Gonzalez-Diaz, H. PTML Multi-Label Algorithms: Models, Software, and Applications. *Curr. Top. Med. Chem.* **2020**, *20*, 2326–2337. [[CrossRef](#)] [[PubMed](#)]
16. Speck-Planche, A.; Cordeiro, M.N.D.S. Multitasking models for quantitative structure-biological effect relationships: Current status and future perspectives to speed up drug discovery. *Expert Opin. Drug Discov.* **2015**, *10*, 245–256. [[CrossRef](#)] [[PubMed](#)]
17. Kleandrova, V.V.; Speck-Planche, A. The urgent need for pan-antiviral agents: From multitarget discovery to multiscale design. *Future Med. Chem.* **2021**, *13*, 5–8. [[CrossRef](#)] [[PubMed](#)]
18. Vasquez-Dominguez, E.; Armijos-Jaramillo, V.D.; Tejera, E.; Gonzalez-Diaz, H. Multioutput Perturbation-Theory Machine Learning (PTML) Model of ChEMBL Data for Antiretroviral Compounds. *Mol. Pharm.* **2019**, *16*, 4200–4212. [[CrossRef](#)]
19. Nocedo-Mena, D.; Cornelio, C.; Camacho-Corona, M.D.R.; Garza-Gonzalez, E.; Waksman de Torres, N.; Arrasate, S.; Sotomayor, N.; Lete, E.; Gonzalez-Diaz, H. Modeling Antibacterial Activity with Machine Learning and Fusion of Chemical Structure Information with Microorganism Metabolic Networks. *J. Chem. Inf. Model.* **2019**, *59*, 1109–1120. [[CrossRef](#)] [[PubMed](#)]
20. Bediaga, H.; Arrasate, S.; Gonzalez-Diaz, H. PTML Combinatorial Model of ChEMBL Compounds Assays for Multiple Types of Cancer. *ACS Comb. Sci.* **2018**, *20*, 621–632. [[CrossRef](#)]
21. Kleandrova, V.V.; Scotti, M.T.; Scotti, L.; Speck-Planche, A. Multi-Target Drug Discovery via PTML Modeling: Applications to the Design of Virtual Dual Inhibitors of CDK4 and HER2. *Curr. Top. Med. Chem.* **2021**, *21*, 661–675. [[CrossRef](#)]
22. Diez-Alarcia, R.; Yanez-Perez, V.; Muneta-Arrate, I.; Arrasate, S.; Lete, E.; Meana, J.J.; Gonzalez-Diaz, H. Big Data Challenges Targeting Proteins in GPCR Signaling Pathways; Combining PTML-ChEMBL Models and [(35)S]GTPgammaS Binding Assays. *ACS Chem. Neurosci.* **2019**, *10*, 4476–4491. [[CrossRef](#)]
23. Kleandrova, V.V.; Speck-Planche, A. PTML Modeling for Alzheimer’s Disease: Design and Prediction of Virtual Multi-Target Inhibitors of GSK3B, HDAC1, and HDAC6. *Curr. Top. Med. Chem.* **2020**, *20*, 1657–1672. [[CrossRef](#)] [[PubMed](#)]
24. Ferreira da Costa, J.; Silva, D.; Caamano, O.; Brea, J.M.; Loza, M.I.; Munteanu, C.R.; Pazos, A.; Garcia-Mera, X.; Gonzalez-Diaz, H. Perturbation Theory/Machine Learning Model of ChEMBL Data for Dopamine Targets: Docking, Synthesis, and Assay of New l-Prolyl-l-leucyl-glycinamide Peptidomimetics. *ACS Chem. Neurosci.* **2018**, *9*, 2572–2587. [[CrossRef](#)] [[PubMed](#)]
25. Abeijon, P.; Garcia-Mera, X.; Caamano, O.; Yanez, M.; Lopez-Castro, E.; Romero-Duran, F.J.; Gonzalez-Diaz, H. Multi-Target Mining of Alzheimer Disease Proteome with Hansch’s QSBR-Perturbation Theory and Experimental-Theoretic Study of New Thiophene Isosters of Rasagiline. *Curr. Drug Targets* **2017**, *18*, 511–521. [[CrossRef](#)]
26. Concu, R.; MN, D.S.C.; Munteanu, C.R.; Gonzalez-Diaz, H. PTML Model of Enzyme Subclasses for Mining the Proteome of Biofuel Producing Microorganisms. *J. Proteome Res.* **2019**, *18*, 2735–2746. [[CrossRef](#)] [[PubMed](#)]
27. Dieguez-Santana, K.; Casanola-Martin, G.M.; Green, J.R.; Rasulev, B.; Gonzalez-Diaz, H. Predicting Metabolic Reaction Networks with Perturbation-Theory Machine Learning (PTML) Models. *Curr. Top. Med. Chem.* **2021**, *21*, 819–827. [[CrossRef](#)] [[PubMed](#)]
28. Santana, R.; Zuluaga, R.; Ganan, P.; Arrasate, S.; Onieva, E.; Gonzalez-Diaz, H. Designing nanoparticle release systems for drug-vitamin cancer co-therapy with multiplicative perturbation-theory machine learning (PTML) models. *Nanoscale* **2019**, *11*, 21811–21823. [[CrossRef](#)] [[PubMed](#)]
29. Santana, R.; Zuluaga, R.; Ganan, P.; Arrasate, S.; Onieva, E.; Gonzalez-Diaz, H. Predicting coated-nanoparticle drug release systems with perturbation-theory machine learning (PTML) models. *Nanoscale* **2020**, *12*, 13471–13483. [[CrossRef](#)]
30. Santana, R.; Zuluaga, R.; Ganan, P.; Arrasate, S.; Onieva, E.; Montemore, M.M.; Gonzalez-Diaz, H. PTML Model for Selection of Nanoparticles, Anticancer Drugs, and Vitamins in the Design of Drug-Vitamin Nanoparticle Release Systems for Cancer Cotherapy. *Mol. Pharm.* **2020**, *17*, 2612–2627. [[CrossRef](#)]
31. Gonzalez-Durruthy, M.; Monserrat, J.M.; Viera de Oliveira, P.; Fagan, S.B.; Werhli, A.V.; Machado, K.; Melo, A.; Gonzalez-Diaz, H.; Concu, R.; MN, D.S.C. Computational MitoTarget Scanning Based on Topological Vacancies of Single-Walled Carbon Nanotubes with the Human Mitochondrial Voltage-Dependent Anion Channel (hVDAC1). *Chem. Res. Toxicol.* **2019**, *32*, 566–577. [[CrossRef](#)] [[PubMed](#)]
32. Kleandrova, V.V.; Luan, F.; Speck-Planche, A.; Cordeiro, M.N.D.S. In silico assessment of the acute toxicity of chemicals: Recent advances and new model for multitasking prediction of toxic effect. *Mini Rev. Med. Chem.* **2015**, *15*, 677–686. [[CrossRef](#)]
33. Martinez-Arzate, S.G.; Tenorio-Borroto, E.; Barbabosa Pliego, A.; Diaz-Albiter, H.M.; Vazquez-Chagoyan, J.C.; Gonzalez-Diaz, H. PTML Model for Proteome Mining of B-Cell Epitopes and Theoretical-Experimental Study of Bm86 Protein Sequences from Colima, Mexico. *J. Proteome Res.* **2017**, *16*, 4093–4103. [[CrossRef](#)]
34. Tenorio-Borroto, E.; Penuelas-Rivas, C.G.; Vasquez-Chagoyan, J.C.; Castanedo, N.; Prado-Prado, F.J.; Garcia-Mera, X.; Gonzalez-Diaz, H. Model for high-throughput screening of drug immunotoxicity—Study of the anti-microbial G1 over peritoneal macrophages using flow cytometry. *Eur. J. Med. Chem.* **2014**, *72*, 206–220. [[CrossRef](#)] [[PubMed](#)]
35. Ghose, A.K.; Viswanadhan, V.N.; Wendoloski, J.J. Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragmental Methods: An Analysis of ALOGP and CLOGP Methods. *J. Phys. Chem. A* **1998**, *102*, 3762–3772. [[CrossRef](#)]

36. Zhu, T.; Cao, S.; Su, P.C.; Patel, R.; Shah, D.; Chokshi, H.B.; Szukala, R.; Johnson, M.E.; Hevener, K.E. Hit identification and optimization in virtual screening: Practical recommendations based on a critical literature analysis. *J. Med. Chem.* **2013**, *56*, 6560–6572. [CrossRef] [PubMed]
37. Makarov, V.; Lechartier, B.; Zhang, M.; Neres, J.; van der Sar, A.M.; Raadsen, S.A.; Hartkoorn, R.C.; Ryabova, O.B.; Vocat, A.; Decosterd, L.A.; et al. Towards a new combination therapy for tuberculosis with next generation benzothiazinones. *EMBO Mol. Med.* **2014**, *6*, 372–383. [CrossRef] [PubMed]
38. Pasca, M.R.; Degiacomi, G.; Ribeiro, A.L.; Zara, F.; De Mori, P.; Heym, B.; Mirrione, M.; Brerra, R.; Pagani, L.; Pucillo, L.; et al. Clinical isolates of *Mycobacterium tuberculosis* in four European hospitals are uniformly susceptible to benzothiazinones. *Antimicrob. Agents Chemother.* **2010**, *54*, 1616–1618. [CrossRef]
39. Xu, J.; Shi, P.Y.; Li, H.; Zhou, J. Broad Spectrum Antiviral Agent Niclosamide and Its Therapeutic Potential. *ACS Infect. Dis.* **2020**, *6*, 909–915. [CrossRef]
40. Fan, X.; Xu, J.; Files, M.; Cirillo, J.D.; Endsley, J.J.; Zhou, J.; Endsley, M.A. Dual activity of niclosamide to suppress replication of integrated HIV-1 and *Mycobacterium tuberculosis* (Beijing). *Tuberculosis (Edinb)* **2019**, *116S*, S28–S33. [CrossRef]
41. Sun, Z.; Zhang, Y. Antituberculosis activity of certain antifungal and antihelminthic drugs. *Tuber. Lung Dis.* **1999**, *79*, 319–320. [CrossRef]
42. Kaushik, A.; Ammerman, N.C.; Tasneen, R.; Story-Roller, E.; Dooley, K.E.; Dorman, S.E.; Nueremberger, E.L.; Lamichhane, G. In vitro and in vivo activity of biapenem against drug-susceptible and rifampicin-resistant *Mycobacterium tuberculosis*. *J. Antimicrob. Chemother.* **2017**, *72*, 2320–2325. [CrossRef]
43. Upton, A.M.; Cho, S.; Yang, T.J.; Kim, Y.; Wang, Y.; Lu, Y.; Wang, B.; Xu, J.; Mdluli, K.; Ma, Z.; et al. In vitro and in vivo activities of the nitroimidazole TBA-354 against *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* **2015**, *59*, 136–144. [CrossRef]
44. Truchon, J.F.; Bayly, C.I. Evaluating virtual screening methods: Good and bad metrics for the “early recognition” problem. *J. Chem. Inf. Model.* **2007**, *47*, 488–508. [CrossRef]
45. Pires, D.E.V.; Ascher, D.B. mycoCSM: Using Graph-Based Signatures to Identify Safe Potent Hits against *Mycobacteria*. *J. Chem. Inf. Model.* **2020**, *60*, 3450–3456. [CrossRef]
46. Gaulton, A.; Bellis, L.J.; Bento, A.P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; et al. ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012**, *40*, D1100–D1107. [CrossRef]
47. Mok, N.Y.; Brenk, R. Mining the ChEMBL database: An efficient chemoinformatics workflow for assembling an ion channel-focused screening library. *J. Chem. Inf. Model.* **2011**, *51*, 2449–2454. [CrossRef]
48. Overington, J. ChEMBL. An interview with John Overington, team leader, chemogenomics at the European Bioinformatics Institute Outstation of the European Molecular Biology Laboratory (EMBL-EBI). Interview by Wendy A. Warr. *J. Comput. Aided Mol. Des.* **2009**, *23*, 195–198. [CrossRef]
49. Ollinger, J.; Kumar, A.; Roberts, D.M.; Bailey, M.A.; Casey, A.; Parish, T. A high-throughput whole cell screen to identify inhibitors of *Mycobacterium tuberculosis*. *PLoS ONE* **2019**, *14*, e0205479. [CrossRef]
50. O’Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, 33. [CrossRef]
51. Valdes-Martini, J.R.; Marrero-Ponce, Y.; Garcia-Jacas, C.R.; Martinez-Mayorga, K.; Barigye, S.J.; Vaz d’Almeida, Y.S.; Pham-The, H.; Perez-Gimenez, F.; Morell, C.A. QuBiLS-MAS, open source multi-platform software for atom- and bond-based topological (2D) and chiral (2.5D) algebraic molecular descriptors computations. *J. Cheminform.* **2017**, *9*, 35. [CrossRef]
52. Valdés-Martini, J.R.; García-Jacas, C.R.; Marrero-Ponce, Y.; Silveira Vaz ‘d Almeida, Y.; Morell, C. QUBILs-MAS: Free Software for Molecular Descriptors Calculator from Quadratic, Bilinear and Linear Maps Based on Graph-Theoretic Electronic-Density Matrices and Atomic Weightings; v1.0; CAMD-BIR Unit, CENDA Registration Number: 2373-2012; Villa Clara, Cuba, 2012. Available online: <http://tomocomd.com/> (accessed on 29 June 2021).
53. Medina Marrero, R.; Marrero-Ponce, Y.; Barigye, S.J.; Echeverria Diaz, Y.; Acevedo-Barrios, R.; Casanola-Martin, G.M.; Garcia Bernal, M.; Torrens, F.; Perez-Gimenez, F. QuBiLS-MAS method in early drug discovery and rational drug identification of antifungal agents. *SAR QSAR Environ. Res.* **2015**, *26*, 943–958. [CrossRef]
54. Marrero-Ponce, Y.; Siverio-Mota, D.; Galvez-Llompart, M.; Recio, M.C.; Giner, R.M.; Garcia-Domenech, R.; Torrens, F.; Aran, V.J.; Cordero-Maldonado, M.L.; Esguera, C.V.; et al. Discovery of novel anti-inflammatory drug-like compounds by aligning in silico and in vivo screening: The nitroindazolinone chemotype. *Eur. J. Med. Chem.* **2011**, *46*, 5736–5753. [CrossRef]
55. Speck-Planche, A.; Cordeiro, M.N.D.S. De novo computational design of compounds virtually displaying potent antibacterial activity and desirable in vitro ADMET profiles. *Med. Chem. Res.* **2017**, *26*, 2345–2356. [CrossRef]
56. Urias, R.W.; Barigye, S.J.; Marrero-Ponce, Y.; Garcia-Jacas, C.R.; Valdes-Martini, J.R.; Perez-Gimenez, F. IMMAN: Free software for information theory-based chemometric analysis. *Mol. Divers.* **2015**, *19*, 305–319. [CrossRef]
57. Godden, J.W.; Bajorath, J. Differential Shannon Entropy as a sensitive measure of differences in database variability of molecular descriptors. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1060–1066. [CrossRef]
58. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106. [CrossRef]
59. Pearson, K. Notes on regression and inheritance in the case of two parents. *Proc. R. Soc. Lond.* **1895**, *58*, 240–242. [CrossRef]
60. Matthews, B.W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta* **1975**, *405*, 442–451. [CrossRef]

61. TIBCO-Software-Inc. *STATISTICA (Data Analysis Software System)*; v13.5.0.17; TIBCO-Software-Inc.: Palo Alto, CA, USA, 2018; Available online: <http://tibco.com> (accessed on 2 July 2021).
62. Sahigara, F.; Mansouri, K.; Ballabio, D.; Mauri, A.; Consonni, V.; Todeschini, R. Comparison of different approaches to define the applicability domain of QSAR models. *Molecules* **2012**, *17*, 4791–4810. [[CrossRef](#)]
63. Netzeva, T.I.; Worth, A.; Aldenberg, T.; Benigni, R.; Cronin, M.T.; Gramatica, P.; Jaworska, J.S.; Kahn, S.; Klopman, G.; Marchant, C.A.; et al. Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. The report and recommendations of ECVAM Workshop 52. *Altern. Lab. Anim.* **2005**, *33*, 155–173. [[CrossRef](#)]
64. Speck-Planche, A. Combining Ensemble Learning with a Fragment-Based Topological Approach to Generate New Molecular Diversity in Drug Discovery: In Silico Design of Hsp90 Inhibitors. *ACS Omega* **2018**, *3*, 14704–14716. [[CrossRef](#)]