Scientific Research Publishing

Computational Chemistry, 2022, 10, 71-96
https://www.scirp.org/journal/cc
ISSN Online: 2332-5984
ISSN Print: 2332-5968

Quantitative Structure-Activity Relationship Study of a Benzimidazole-Derived Series Inhibiting *Mycobacterium tuberculosis H37Rv*

Georges Stéphane Dembélé[1,2], Mamadou Guy-Richard Koné[1,2*], Fandia Konate[1], Doh Soro[1], Nahossé Ziao[1,2]

[1]Laboratoire de Thermodynamique et de Physico-Chimie du Milieu, UFR SFA, Université Nangui Abrogoua, Abidjan, Côte-d'Ivoire
[2]Groupe Ivoirien de Recherches en Modélisation des Maladies (GIR2M), Université Nangui Abrogoua, Abidjan, Côte-d'Ivoire
Email: *guyrichardkone@gmail.com

How to cite this paper: Dembélé, G.S., Koné, M.G.-R., Konate, F., Soro, D. and Ziao, N. (2022) Quantitative Structure-Activity Relationship Study of a Benzimidazole-Derived Series Inhibiting *Mycobacterium tuberculosis H37Rv*. *Computational Chemistry*, **10**, 71-96.
https://doi.org/10.4236/cc.2022.102004

Received: February 16, 2022
Accepted: April 9, 2022
Published: April 12, 2022

Abstract

This work was carried out on a series of twenty-two (22) benzimidazole derivatives with inhibitory activities against *Mycobacterium tuberculosis H37Rv* by applying the Quantitative Structure-Activity Relationship (QSAR) method. The molecules were optimized at the level DFT/B3LYP/6−31 + G (d, p), to obtain the molecular descriptors. We used three statistical learning tools namely, the linear multiple regression (LMR) method, the nonlinear regression (NLMR) and the artificial neural network (ANN) method. These methods allowed us to obtain three (3) quantitative models from the quantum descriptors that are, chemical potential (μ), polarizability (α), bond length l (C = N), and lipophilicity. These models showed good statistical performance. Among these, the ANN has a significantly better predictive ability R^2 = 0.9995; RMSE = 0.0149; F = 31879.0548. The external validation tests verify all the criteria of Tropsha *et al.* and Roy *et al.* Also, the internal validation tests show that the model has a very satisfactory internal predictive character and can be considered as robust. Moreover, the applicability range of this model determined from the levers shows that a prediction of the pMIC of the new benzimidazole derivatives is acceptable when its lever value is lower than 1.

Keywords

Mycobacterium tuberculosis H37Rv, Benzimidazole Derivatives, QSAR, ANN, Applicability Domain

1. Introduction

Tuberculosis is an infectious and contagious disease caused by Koch's bacillus (strains of the *Mycobacterium tuberculosis* complex). This infectious agent is transmitted by air, via droplets containing the bacteria and expectorated by the cough of the patients. Tuberculosis is present in all regions of the world [1]. In 2019, the WHO Region with the highest number of new TB cases was Southeast Asia (44% of all new cases), followed by the African Region (25%) and the Western Pacific Region (18%). In 2019, 87% of all new cases occurred in the 30 countries with the highest TB burden. Two-thirds of new cases were concentrated in eight countries: India, Indonesia, China, the Philippines, Pakistan, Nigeria, Bangladesh, and South Africa [2]. In Ivory Coast, data from the World Health Organization (WHO) 2020 report indicate that the incidence of TB is 137 cases per 100,000 populations. The number of notified cases was 21,498 in 2019 and 19,976 in 2020 [3]. However, TB is a treatable and curable disease. Patients with drug-susceptible active tuberculosis receive a standard 6-month course of four antimicrobial drugs and are given information and support by a trained health worker or volunteer. Despite this treatment, resistance to isoniazid and rifampicin, the two most effective first-line anti-TB drugs, has been observed [2]. In 2019, 206,030 cases of multidrug-resistant tuberculosis or rifampin-resistant tuberculosis were detected and reported worldwide, a 10% increase from 186,883 cases in 2018. The design of anti-tuberculosis antibiotics above any resistance of *Mycobacterium tuberculosis* remains a challenge for the scientific community. It is in this context that, pharmacochemists are interested in the research of compounds with pharmacological activities of pharmaceutical interest [4]. Raynaud *et al.* [5] have shown that benzimidazole derivatives possess activity against *Mycobacterium tuberculosis H37rv*. Benzimidazole derivatives are associated with a wide range of biological activities. They have anticancer [6], anti-VIH [7], antibacterial [8], anti-inflammatory [9], antihistamine, antioxidant [10], antihypertensive [11] activities etc.

Quantitative structure-activity relationship (QSAR) is a technique that consists in relating the molecular structure to a well-defined parameter such as biological activity. This method allows to reduce the excessive number of experiments, sometimes long, dangerous and costly in terms of time and money [12] [13]. The overall objective of this work is to develop reliable models to explain and predict the MIC (minimum inhibitory concentration in μg/ml) antituberculosis activity of a series of twenty (22) benzimidazole derivatives (**Figure 1**).

2. Materials and Methods

2.1. Computational Theory Level

In order to predict the antitubercular activity of benzimidazole derivatives, descriptors were determined by quantum chemical calculations using Gaussian 09 [14]. DFT methods are known for their ability to provide a multitude of molecular properties in QSAR studies [15] [16] [17]. These increase the predictive

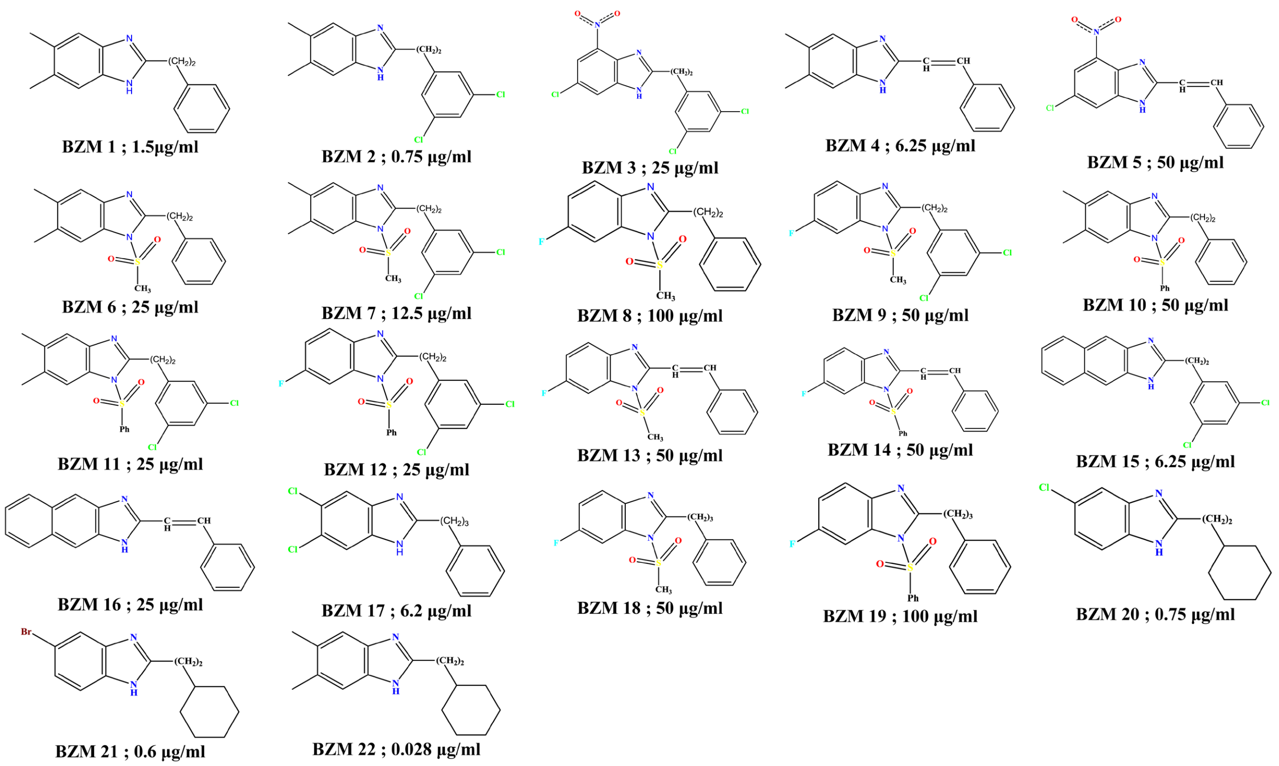

Figure 1. Molecular structure, code and inhibitory concentration (MIC) of the twenty-two (22) benzimidazole derivatives.

capability of QSAR models while shortening the computational time and cost implication in the design of new drugs [18] [19]. The twenty-two (22) compounds used to conduct this work have minimum inhibitory concentration (MIC) values that range from 0.08 to 100 µg/mL. Minimum inhibitory concentration (MIC) is an indicator of the effectiveness of a given compound in inhibiting a specific biological or biochemical function. Concentration values are usually expressed as the inverse of the logarithm of activity based on decimal places $\left(-\log_{10}(C)\right)$ to obtain better mathematical values when structures are biologically active [20] [21]. The anti-tuberculosis activity will be expressed by the potential inhibitory concentration pMIC defined by Equation (1):

$$\text{pMIC} = -\log_{10}\left(\frac{\text{MIC}}{M} * 10^{-3}\right) \quad (1)$$

where MIC, the minimum inhibitory concentration in µg/mL.

The data modeling was developed using three statistical learning methods. These are the Linear Multiple Regression (LMR) and Non-Linear Multiple Regression (NLMR) methods that are integrated in Excel [22] and XLSTAT [23]. The last method is that of artificial neurons which is included in the JMP Pro software [24].

2.2. Molecular Descriptors Used

In the development of our QSAR model, quantum descriptors have been calculated. In particular, chemical potential (μ), polarizability (a), bond length (l(C =

N)) and lipophilicity (log$P$). The chemical potential μp measures the tendency of the electron cloud to escape from the molecule. Also, the larger the value of this parameter, the greater the reactivity of the molecule with a nucleophile.

$$\mu = -\frac{I + AE}{2} \tag{2}$$

With:

$$I = -E_{HO} \tag{3}$$

$$AE = -E_{BV} \tag{4}$$

The polarizability designates a phenomenon caused by the moment of the electric charges of the atom. A molecule placed in an electric field E undergoes a deformation and acquires an induced dipole electric moment proportional to the field E, the polarizabilities are expressed in Å^3. They have the dimension of a volume. The atomic polarizability increases with the size of the atoms [25].

$$\alpha = \varepsilon_0 \mu E \tag{5}$$

where:

α: Polarizability coefficient;

ε_0 : Dielectric constant;

μ: Induced dipole electric moment.

The geometric descriptor used is the bond length l (C = N) in Armstrong (Å) (**Figure 2**). This descriptor is illustrated by the figure below around the benzimidazole ring.

Lipophilicity reflects the ability of a molecule to adhere to a lipidic environment, oil, cell membrane, lipidic solvent, etc. [26]. This physico-chemical descriptor is generally evaluated by the distribution of the molecule, neutral, soluble, between water and another immiscible solvent: most often n-octanol (or octan-1-ol) [26] [27] [28] [29]. This parameter is estimated from the log$P$ value. The log$P$ is equal to the logarithm of the ratio of the concentrations of the test substance in octanol and in water log$P$ = log(C_{oct}/C_{Water}). Indeed, if the log$P$ is positive and very high, it expresses the fact that the molecule considered is much more soluble in octanol than in water, which reflects its lipophilic character, and conversely, if the log$P$ is negative it means that the molecule considered is hydrophilic. A zero log$P$ means that the molecule is as soluble in one solvent as in the other. In this work, the software Chemsketch [30] allowed us to determine the values of log$P$. In practice we express the lipophilicity by the decimal logarithm of the partition coefficient log$P$. Thus:

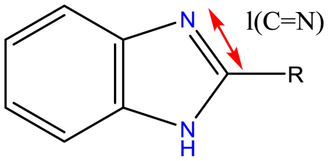

Figure 2. Geometric descriptor of the benzimidazole derivatives used: the bond length l (C = N) in Armstrong (Å).

If $\log P > 0$; then $P > 1$, the molecule is lipophilic. It is soluble in the lipidic phase. It is then not polar.

If $\log P < 0$; then $P < 1$, the molecule is hydrophilic. It is soluble in water. It is then polar.

2.3. Estimation of the Predictive Capacity of a QSAR Model

The quality of a model is determined based on different statistical criteria of analysis including the coefficient of determination R^2, standard deviation (S) or Root Mean Square Error (RMSE), cross-validation correlation coefficients Q^2_{CV} and Fischer F. R^2, S and F refer to the fit of the simulated and experimental values. They represent the predictive capacity within the limits of the model, and allow us to estimate the accuracy of the values calculated on the test set [31] [32]. As for the cross-validation coefficient Q^2_{CV}, it provides information on the predictive character of the model. This predictive potential is said to be "internal" because it is calculated from the structures used to build the model. The coefficient of determination R^2 gives an evaluation of the dispersion of the theoretical values around the experimental values. The quality of the modeling is better when the points are close to the fit line [33]. The fit of the points to this line can be evaluated by the coefficient of determination.

$$R^2 = 1 - \frac{\sum\left(y_{i,exp} - \hat{y}_{i,theo}\right)^2}{\sum\left(y_{i,exp} - \bar{y}_{i,exp}\right)^2} \tag{6}$$

where:

$y_{i,exp}$: Experimental value of anti-tuberculosis activity;

$\hat{y}_{i,theo}$: Theoretical value of anti-tuberculosis activity;

$\bar{y}_{i,exp}$: Mean value of experimental values of anti-tuberculosis activity.

The closer the R^2 value is to 1, the more the theoretical and experimental values are correlated

The Root Mean Square Error RMSE is another statistical indicator used. It allows to evaluate the reliability and the precision of a model:

$$\text{RMSE} = \sqrt{\frac{\sum\left(y_{i,exp} - y_{i,theo}\right)^2}{n-k-1}} \tag{7}$$

The Fisher **F** test is also used to measure the level of statistical significance of the model, *i.e.* the quality of the choice of descriptors constituting the model

$$F = \frac{\sum\left(y_{i,theo} - y_{i,exp}\right)^2}{\sum\left(y_{i,exp} - y_{i,theo}\right)^2} * \frac{n-k-1}{k} \tag{8}$$

The coefficient of determination of the cross-validation Q^2_{CV} allows to evaluate the accuracy of the prediction on the training set. It is calculated using the following relation:

$$Q^2_{cv} = \frac{\sum\left(y_{i,theo} - \bar{y}_{i,exp}\right)^2 - \sum\left(y_{i,theo} - y_{i,exp}\right)^2}{\sum\left(y_{i,theo} - \bar{y}_{i,exp}\right)^2} \tag{9}$$

2.4. Acceptance Criteria of a Model

The performance of a mathematical model, according Eriksson *et al.* [34], is characterized by a value of $Q_{cv}^2 > 0.5$ for a satisfactory model, when for the excellent model $Q_{cv}^2 > 0.9$. According to these authors, given a set of tests, a model will perform well if the acceptance criterion $R^2 - Q_{cv}^2 < 0.3$ is met.

According to Tropsha *et al.* [35] [36] [37], or the external validation set, the predictive power of a model can be obtained from five criteria. These criteria are the following:

$$
\begin{aligned}
&1)\quad R_{Test}^2 > 0.7, \\
&2)\quad Q_{Cv\,Test}^2 > 0.6, \\
&3)\quad \left|R_{Test}^2 - R_0^2\right| \le 0.3, \\
&4)\quad \frac{\left|R_{Test}^2 - R_0^2\right|}{R_{Test}^2} < 0.1 \quad \text{et} \quad 0.85 \le k \le 1.15, \\
&5)\quad \frac{\left|R_{Test}^2 - R_0'^2\right|}{R_{Test}^2} < 0.1 \quad \text{et} \quad 0.85 \le k' \le 1.15.
\end{aligned}
\tag{10}
$$

In addition, Roy and Roy [38], have refined the prediction method of a QSAR model. They have developed quantities r_m^2 and Δr_m^2, called metric values. r_m^2 determines the closeness between the observed activity and the prediction. The metric values r_m^2 and Δr_m^2 are calculated from the observed and predicted activities. Currently, these two different variants r_m^2 and Δr_m^2 can be calculated for the test set (internal validation) or for the test set (external validation). A QSAR model is acceptable to these authors, if both of these criteria are met.

$$\overline{r_m^2} = \frac{r_m^2 + r_m'^2}{2} > 0.5 \tag{11}$$

$$\Delta r_m^2 = \left|r_m^2 - r_m'^2\right| < 0.2$$

where $r_m^2 = r^2 * \left(1 - \sqrt{\left(r^2 - r_0^2\right)}\right)$ et $r_m'^2 = r^2 * \left(1 - \sqrt{\left(r^2 - r_0'^2\right)}\right)$.

2.5. Statistical Analysis

2.5.1. Linear and Non-Linear Multiple Regressions (LMR and NLMR)

The statistical method of multiple linear regression (MRL) is used to examine the relationship between a dependent variable (Property) and various independent variables (Descriptors). This statistical approach limits the differences between the actual and predicted values. It was also used to select the descriptors used as input parameters in the multiple non-linear regression (NLMR). As for the NLMR analysis, it is also used to refine the structure-property relationship in order to quantitatively assess the property. It is the most common tool for studying multidimensional data. It is based on the following pre-programmed functions of XLSTAT:

$$y = a + \left(bx_1 + cx_2 + dx_3 + ex_4\right) + \left(fx_{12} + gx_{22} + hx_{32} + ix_{42}\right) \tag{12}$$

where $a, b, c, d, \cdots$ represent the parameters and, $x_1, x_2, x_3, x_4, \cdots$ represent the variables.

2.5.2. Artificial Neural Network (ANN)

Artificial neurons are inspired by the human biological neuron. As such, they are made up of cells or neurons linked together by connections that allow them to send and receive signals from other cells. These networks are mathematical models composed of several neurons, arranged in different layers. In principle, the network is composed of three layers: an input layer, a hidden layer and an output layer, connected by a complex network. [39] [40]. The most commonly used networks are multilayer perceptrons (MLPs) whose neurons are usually divided into layers [41]. In this paper, the artificial neural network was made from the 4-3-1 multilayer perceptron network, *i.e.*, the network consists of five (4) neurons in the input layer, three (3) neurons in the hidden layer and one (1) neuron in the output layer. The output layer consists of a sigmoid function. The architecture of the ANN models used is described in **Figure 3** below.

2.6. Area of Applicability

The domain of applicability of a QSAR model is the physicochemical, structural or biological field in which the model equation can be used to make predictions about new drugs [42]. It corresponds to the area of the chemical space including the compounds of the training set and similar compounds, which are close in this same space [43]. In particular, the model, which is built on the basis of a limited number of compounds, by relevant descriptors cannot be a universal tool to predict the activity of any other molecule with certainty. It appears necessary, to determine the DA of any QSAR model. This is recommended by the Organization (OECD) in the development of a QSAR model of economic cooperation and development [44]. There are different methods to establish the domain of applicability of a model [43]. Among these, the method used in this work is the leverage approach. This method is based on the variation of the standardized residuals of the dependent variable with the distance between the values of the descriptors and their mean, called leverage [45]. The h_{ii} are the diagonal elements of a matrix H called hat matrix. H is the projection matrix of the experimental values of the explained variable Y_{expe} in the space of the predicted values of the explained variable Y_{pred} such that:

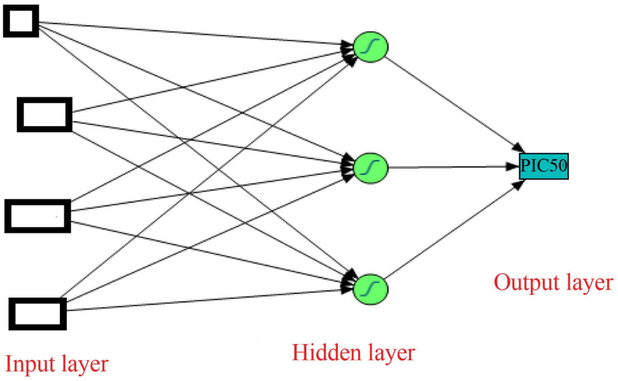

Figure 3. Diagram of the structure of a multilayer perceptron.

$$Y_{pred} = HY_{expe} \tag{13}$$

H is defined by the expression (12):

$$H = X\left(X^{t}X\right)^{-1}X^{t} \tag{14}$$

The field of applicability is delimited by a threshold value of the lever noted h^*. In general, it is fixed at $3\frac{p+1}{n}$, where n is the number of compounds in the training set, and p is the number of descriptors in the model [46] [47]. For standardized residuals, the two limit values generally used are ±3σ, σ being the standard deviation of the experimental values of the quantity to be explained [48]: it is "the rule of the three sigmas" [49].

3. Results and Discussion

In our QSAR study, we used a series of twenty-two (22) benzimidazole derivatives. These compounds were synthesized and tested on *Mycobacterium tuberculosis H*37*Rv*. The compounds were split into two groups, fifteen (15) were used for the learning set and seven (7) for the validation set. This part of the work will concern the modeling of the antitubercular activity of the benzimidazole derivatives based on the descriptors presented in **Table 1**. The values of the descriptors as well as those of the experimental biological activities of the molecules are recorded in **Table 1**.

3.1. Interdependence of Descriptors

In order to better understand the interdependence of the descriptors used, we present the values of the partial correlation coefficients aij of these descriptors in **Table 2**.

The partial correlation coefficients a_{ij} contained in **Table 2** between the pairs of descriptors (μ, α), (μ, I(C = N)), (μ, log$P$), (α, I(C = N)), (α, log$P$), (I(C = N), log$P$) are less than 0.7 ($a_{ij} < 0.7$). This demonstrates the independence of the descriptors used to develop the models.

3.2. Modeling of *Mycobacterium tuberculosis H*37*Rv* Activity

In the model formula, the negative or positive sign of the coefficient of a descriptor reflects the effect of proportionality between the evolution of the inhibitory concentration MIC and this physicochemical parameter of the regression equation. Thus, the negative sign indicates that when the value of the descriptor is high, the MIC inhibitory concentration decreases, whereas the positive sign expresses the opposite effect. In this work, three statistical analysis tools were used: Multiple Linear Regression (MLR), Multiple Nonlinear Regression (MNLR) and Artificial Neural Network (ANN).

3.2.1. Multiple Linear Regression (MLR)

The equation of the QSAR model is presented below. The statistical indicators are given in **Table 3**.

Table 1. Physicochemical descriptors and experimental pMICs of the learning and validation sets.

Molecules	μ (eV)	α (Å^3)	l(C = N) (Å)	log$P$	pMIC
			Training Set		
BZM2	−3.5455	248.1263	1.3126	5.8700	5.6290
BZM3	−4.9862	255.3323	1.3138	4.9700	4.1710
BZM4	−3.7021	267.5480	1.3218	5.2300	4.5991
BZM6	−3.6587	256.3247	1.3018	4.2000	4.1185
BZM7	−3.8331	283.6173	1.3017	5.4000	4.5022
BZM9	−4.0971	255.4973	1.3011	4.5400	3.8890
BZM10	−4.0244	310.8677	1.3019	5.9500	3.8927
BZM11	−4.1378	338.7730	1.3020	7.1600	4.2642
BZM13	−4.1931	269.7247	1.3099	3.8700	3.8012
BZM14	−4.1352	322.5207	1.3102	5.6300	3.8790
BZM15	−3.5924	280.8597	1.3085	6.1800	4.7372
BZM16	−3.9000	308.1380	1.3200	5.5400	4.0340
BZM19	−4.2359	293.4540	1.3035	5.5200	3.5960
BZM21	−3.6102	214.6073	1.3143	5.6200	5.7093
BZM22	−3.1385	218.2873	1.3140	5.6200	6.5058
			Test Set		
BZM1	−3.2413	220.4227	1.3132	4.6600	5.2224
BZM5	−4.5969	279.8017	1.3245	4.7300	3.7777
BZM8	−3.9554	227.5893	1.3012	3.3300	3.5029
BZM12	−4.3856	309.2930	1.3013	6.2900	4.2546
BZM17	−3.7969	230.3647	1.3142	5.0500	4.6922
BZM18	−3.9513	239.7053	1.3027	3.7600	3.8227
BZM20	−3.6134	204.8540	1.3147	5.1900	5.5445

Table 2. Correlation matrix between the different physico-chemical descriptors.

Variables	μ	α	l(C = N)	log$P$
μ	**1.0000**			
α	−0.3950	**1.0000**		
l(C = N)	0.1685	−0.3032	**1.0000**	−0.0240
log$P$	0.1715	0.4270	−0.0240	**1.0000**

Table 3. Statistical analysis ratio of the minimum inhibitory concentration (pMIC) potential of benzimidazole derivatives of RML model.

Number of observations N	15
Coefficient of determination R^2	0.9204
Standard deviation RMSE	0.2661
Fischer test F	150.311
Cross-validation correlation coefficient Q^2_{CV}	0.9204
Confidence level α	>95%

$$\text{pMIC}^{\text{th}} = 0.59453 * \mu - 0.01829 * \alpha + 6.69316 * l(\text{C=N}) + 0.56907 * \log P$$

The negative sign of the coefficient of polarizability (α) reflects that antitubercular activity will be improved for low values of this descriptor. In contrast, the positive sign of the coefficient of chemical potential (μ), C = N bond length and lipophilicity (Log$P$) indicates that high values of these descriptors will improve anti-tuberculosis activity.

The coefficient of determination (R^2 = 0.9204), shows that the predicted pMIC values contain 92.04% of the experimental values. The Fisher test value (F = 150.311) is very high compared to the critical value, from the Fisher table Fcr = 2.96 [50]. This value 150.311 of the Fisher test, higher than the critical value, shows that the error committed is lower than what the model explains [50]. The standard deviation (RMSE = 0.2661) expresses the small deviation of the predicted values from the experimental mean. This model presents a correlation coefficient of the cross-validation Q^2_{cv} equal to $Q^2_{cv} = 0.9204$. This value, higher than 0.9, reflects a so-called excellent model according to Erikson *et al.* [34]. This model is acceptable because it is in agreement with the acceptance criteria of these authors $R^2 - Q^2_{cv} = 0.9204 - 0.9204 = 0.000 < 0.3$. All these statistical indicators show that the model developed explains the anti-tuberculosis activity in a statistically significant and satisfactory manner. These different results are confirmed by the regression plot of the MLR model presenting the theoretical antimalarial activity as a function of the experimental activity represented in **Figure 4**.

The regression curve of the MLR model shows that all points are around the regression line. This result indicates that there is a small difference (RMSE = 0.2661) between the values of pMIC$^{\text{exp}}$ and pMIC$^{\text{th}}$, thus a good similarity in these values. This similarity is illustrated in **Figure 5**.

1) Internal validation

Internal validation of the MLR model was performed using the Leave One Out (LOO) procedure and the randomization test.

a) Leave-One-Out (LOO) procedure

The leave-one-out (LOO) cross-validation procedure was applied on the 15 molecules of the training set. The results obtained are presented in **Table 4**.

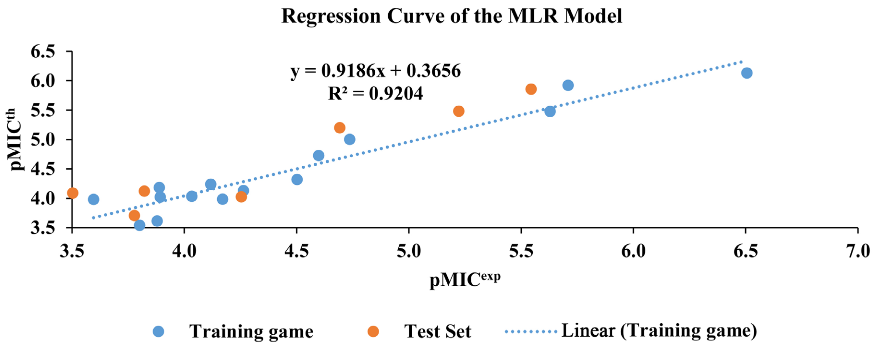

Figure 4. The regression line of the MLR model.

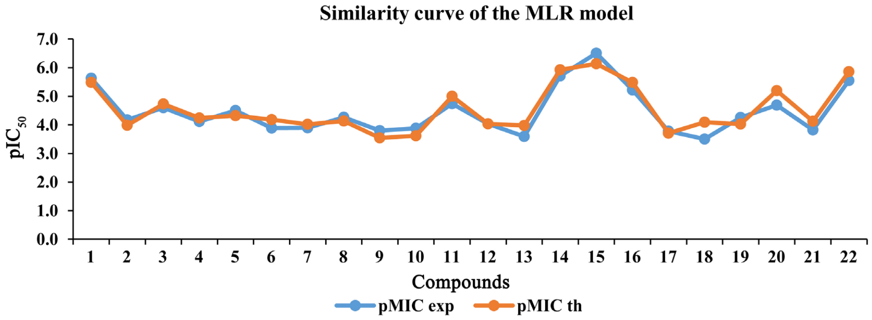

Figure 5. Similarity curve of experimental and predicted values of the MLR model.

Table 4. Statistical parameters of the leave-one-out (LOO) cross-validation of the MLR model.

MOLECULES	pMIC^{exp}	pMIC^{pred}	R^2	RMSE	F	Q^2_{LOO}
BZM2	5.6290	5.4476	0.9104	0.2741	121.9910	
BZM3	4.1710	3.3711	0.9347	0.2514	171.7805	
BZM4	4.5991	4.7457	0.9223	0.2756	142.3927	
BZM6	4.1185	4.2946	0.9214	0.2752	140.7552	
BZM7	4.5022	4.3021	0.9241	0.2724	146.2007	
BZM9	3.8890	4.2365	0.9281	0.2599	154.9143	
BZM10	3.8927	4.0446	0.9193	0.2755	136.6950	**0.7936**
BZM11	4.2642	4.0091	0.9234	0.2729	144.7556	
BZM13	3.8012	3.3932	0.9274	0.2595	153.3583	
BZM14	3.8790	3.5492	0.9262	0.2631	150.7110	
BZM15	4.7372	5.0507	0.9285	0.2636	155.8035	
BZM16	4.0340	4.0333	0.9186	0.2790	135.4085	
BZM19	3.5960	4.0320	0.9316	0.2472	163.3645	
BZM21	5.7093	6.0453	0.9137	0.2658	127.1142	
BZM22	6.5058	5.8983	0.8981	0.2351	105.7664	
	Averages		**0.9219**	**0.2647**	**143.4008**	

The results show that the models constructed, after the removal of one of the compounds from the training set (first column of the table), have statistical parameters (R^2 and RMSE) of the same order as those of the initial model, overall. The average values of these parameters are $R^2 = 0.9219$, RMSE = 0.2647. We find values almost identical to those of the initial model. The cross-correlation coefficient Q_{cv}^2, is equal to 0.7936. This value is higher than the minimum required value of 0.5 according to Tropsha *et al.* [51] [52]. In addition, we note that $R^2 - Q_{cv}^2 = 0.1283 < 0.3$ [53]. All this shows that the RML model has a very satisfactory internal predictive character and can be considered as robust [54].

b) Randomization test

The randomization test of the MLR model was performed on the molecules of the training set by randomly permuting the values of the activities while keeping the descriptors for model building. We stopped at ten (10) iterations. The randomized coefficients of determination (R_r^2) for each iteration are listed in **Table 5**.

From the values in **Table 5**, the value of Roy's parameter ($R_p^2 = 0.5593$) was determined. This value ($R_p^2 = 0.5593$) is lower than the coefficient of determination of the model (0.9204). These different results show that the is not due to chance and can be considered as robust.

2) External validation

The external validation of the RML model was performed on the molecules of the validation set (**Table 1**) using the Tropsha criteria [51] [52] and Roy [38]. The Tropsha and Roy criteria checks are recorded in **Table 6** and **Table 7** respectively.

Table 5. Randomized coefficients of determination (R_r^2) of the ten (10) iterations.

Iteration	1	2	3	4	5	6	7	8	9	10
R_r^2	0.6765	0.6646	0.5768	0.5524	0.5891	0.4757	0.5593	0.5109	0.3665	0.5384

Table 6. Tropsha criteria checks of the external validation set of the MLR model.

Statistical parameters	Tropsha criteria [35] [36] [37]	
R^2	>0.7	**0.881**
Q_{CV}^2	>0.6	**0.746**
$\left\|R^2 - R_0^2\right\|$	≤0.3	**0.001**
$\frac{\left\|R^2 - R_0^2\right\|}{R^2}$	<0.1	**0.001**
k	$0.85 \le k \le 1.15$	**1.054**
$\frac{\left\|R^2 - R_0'^2\right\|}{R^2}$	<0.1	**0.011**
k'	$0.85 \le k' \le 1.15$	**0.946**

Table 7. Roy criteria checks of the external validation set of the MLR model.

Indicators	r_m^2	$r_m'^2$	$\overline{r_m^2} = \frac{r_m^2 + r_m'^2}{2}$	$\Delta r_m^2 = \left| r_m^2 - r_m'^2 \right|$
Value	**0.857**	**0.795**	**0.826**	**0.062**

The values in Table 6 show that all Tropsha criteria are met, so the model is acceptable for predicting the antitubercular activity of benzimidazole derivatives.

The analysis in Table 7 shows that the r_m^2 is greater than 0.5 and the Δr_m^2 is less than 0.2. This result reflects that the model meets Roy's criteria. We can therefore affirm that the model is robust and has a good predictive power.

3.2.2. Non-Linear Multiple Regression NLMR

The equation of the QSAR model is presented below. The statistical indicators are given in Table 8.

$$\begin{aligned} \text{pMIC}^{\text{th}} &= -2682 + 5.81274 * \mu - 0.04293 * \alpha + 4118 * l(\text{C=N}) \\ &+ 0.01167 * \log P + 0.61675 * \mu^2 + 0.00006 * \alpha^2 \\ &- 1566 * l(\text{C=N})^2 + 0.03466 * \log P^2 \end{aligned}$$

The coefficient of determination ($R^2 = 0.9648$), shows that the predicted pMIC values contain 96.48% of the experimental values. The Fisher test value (F = 356.324) is very high compared to the critical value, from the Fisher table Fcr = 2.96 [50]. This value 356.324 of the Fisher test, higher than the critical value, shows that the error committed is lower than what the model explains [50]. The standard deviation (RMSE = 0.2396) expresses the small deviation between the predicted values and the experimental mean. This model presents a correlation coefficient of the cross-validation Q_{cv}^2 equal to $Q_{cv}^2 = 0.9648$. This value, higher than 0.9, reflects a so-called excellent model according to Erikson *et al.* [34]. This model is acceptable because it is in agreement with the acceptance criteria of these authors $R^2 - Q_{cv}^2 = 0.9648 - 0.9248 = 0.000 < 0.3$. All these statistical indicators show that the model developed explains the TB activity in a statistically significant and satisfactory way. These different results are confirmed by the regression plot of the NLMR model presenting the theoretical anti-tuberculosis activity as a function of the experimental activity represented in Figure 6.

The regression curve of the RML model shows that all points are around the regression line. This result indicates that there is a small difference (RMSE = 0.2396) between the values of pMICexp and $\text{pMIC}^{\text{th,}}$ thus a good similarity in these values. This similarity is illustrated in Figure 7.

1) Internal validation

a) Randomization test

The NLMR model randomization test was performed on the molecules in the training set by randomly permuting the activity values while retaining the descriptors for model building. We stopped at ten (10) iterations. The randomized coefficients of determination (R_r^2) for each iteration are listed in Table 9.

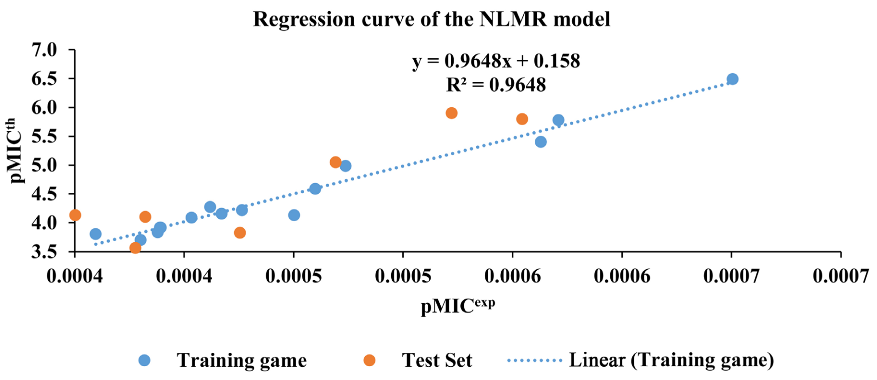

Figure 6. The regression line of the NLMR model.

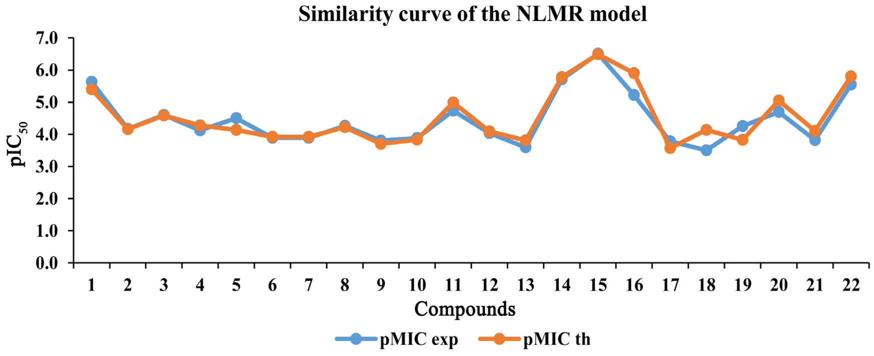

Figure 7. Similarity curve of experimental and predicted values of NLMR model.

Table 8. Statistical analysis ratio of potential inhibitory concentration (pMIC) of benzimidazole derivatives from the NLMR model.

Number of observations N	15
Coefficient of determination R^2	0.9648
Standard deviation RMSE	0.2396
Fischer test F	356.3245
Cross-validation correlation coefficient Q_{CV}^2	0.9648
Confidence level α	>95%

Table 9. Randomized coefficients of determination (R_r^2) of the ten (10) iterations.

Iteration	1	2	3	4	5	6	7	8	9	10
R_r^2	0.7125	0.6809	0.5593	0.7279	0.7767	0.7009	0.6906	0.8160	0.7013	0.4854

From the values in Table 9, the value of Roy's parameter ($R_p^2 = 0.5102$) was determined. This value ($R_p^2 = 0.5102$) is lower than the coefficient of determination of the model (0.9648). These different results show that the is not due to chance and can be considered as robust.

2) External Validation

External validation of the NLMR model was performed on the molecules in the validation set (Table 1) using the Tropsha [51] [52] and Roy [38]. The Tropsha and Roy criteria checks are recorded in Table 10 and Table 11, respectively.

The values in Table 10 show that all Tropsha criteria are met, so the model is acceptable for predicting the antitubercular activity of benzimidazole derivatives.

The analysis in Table 11 shows that r_m^2 is greater than 0.5 and the Δr_m^2 is less than 0.2. This result reflects that the model meets Roy's criteria. We can therefore affirm that the model is robust and has good predictive power.

3.2.3. Contribution of the Descriptors of the MLR and NLMR Models

The study of the relative contribution of the descriptors in predicting the antitubercular activity of benzimidazole derivatives was performed. The different contributions are presented in the pie chart in Figure 8.

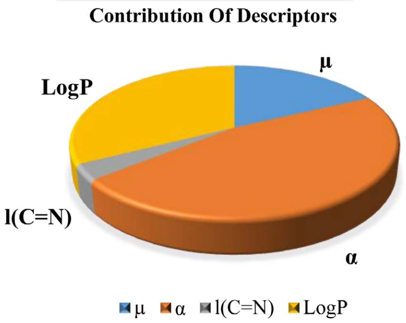

Figure 8. Contribution of descriptors in the models.

Table 10. Tropsha criteria checks of the external validation set of the NLMR model.

Statistical parameters	Tropsha criteria [35] [36] [37]	
R^2	>0.7	**0.824**
Q^2_{CV}	>0.6	**0.623**
$\left|R^2 - R_0^2\right|$	≤0.3	**0.003**
$\frac{\left|R^2 - R_0^2\right|}{R^2}$	<0.1	**0.003**
k	$0.85 \le k \le 1.15$	**1.053**
$\frac{\left|R^2 - R_0'^2\right|}{R^2}$	<0.1	**0.079**
k'	$0.85 \le k' \le 1.15$	**0.944**

Table 11. Roy criteria checks of the external test set of the NLMR model.

Indicators	r_m^2	$r_m'^2$	$\overline{r_m^2} = \frac{r_m^2 + r_m'^2}{2}$	$\Delta r_m^2 = \left|r_m^2 - r_m'^2\right|$
Value	**0.782**	**0.613**	**0.698**	**0.169**

From **Figure 5**, we can see that polarizability (α) is the descriptor with the highest contribution compared to the other descriptors. Thus, polarizability (α) is the priority descriptor in predicting anti-tuberculosis activity.

3.2.4. Area of Applicability of the MLR and NLMR Models

The applicability domain of the MLR and NLMR models was determined by the leverage method. The h_{ii}-lever values of the molecules in the training set calculated from the MINTAB software are listed in **Table 12**.

The values of the h_{ii} levers in **Table 12** and the standardized residues of the molecules were used to plot the graph of standardized residues versus h_{ii} levers in **Figure 9**.

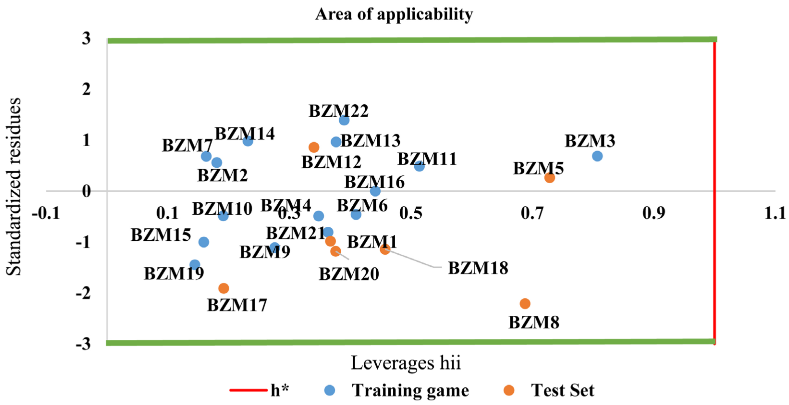

Figure 9. Graph of standardized residuals of antituberculosis activity according to the levers of the MLR and NLMR models.

Table 12. Lever values of the molecules in the training set.

Molecules	h_{ii}
BZM2	0.1805
BZM3	0.8071
BZM4	0.3483
BZM6	0.4100
BZM7	0.1632
BZM9	0.2761
BZM10	0.1914
BZM11	0.5144
BZM13	0.3772
BZM14	0.2320
BZM15	0.1594
BZM16	0.4418
BZM19	0.1446
BZM21	0.3637
BZM22	0.3904

For the 15 molecules of the training set and the 4 descriptors of the model, the threshold value of the h^* levers is 1. The extreme values of the standardized residuals are ±3 according to the "three sigma rule" [49]. These different values delimit the field of applicability [55] of the model as indicated on the graph in Figure 6. Figure 6 shows us that all the molecules have levers lower than the threshold lever ($h^* = 1$) and values of the standardized residues between +3 and −3. This result thus translates that all the molecules belong to the applicability domain.

3.2.5. Artificial Neural Network (ANN)

The values of the descriptors as well as those of the experimental biological activities of the molecules used for the development of the ANN model are listed in Table 13.

Table 13. Experimental physicochemical and pMIC descriptors of the ANN model training and validation sets.

Molécules	μ	α	$I(C = N)$	$\log P$	pMIC
			Training Set		
BZM1	−3.2413	220.4227	1.3132	4.6600	5.2224
BZM2	−3.5455	248.1263	1.3126	5.8700	5.6290
BZM3	−4.9862	255.3323	1.3138	4.9700	4.1710
BZM4	−3.7021	267.5480	1.3218	5.2300	4.5991
BZM5	−4.5969	279.8017	1.3245	4.7300	3.7777
BZM8	−3.9554	227.5893	1.3012	3.3300	3.5029
BZM9	−4.0971	255.4973	1.3011	4.5400	3.8890
BZM10	−4.0244	310.8677	1.3019	5.9500	3.8927
BZM12	−4.3856	309.2930	1.3013	6.2900	4.2546
BZM13	−4.1931	269.7247	1.3099	3.8700	3.8012
BZM15	−3.5924	280.8597	1.3085	6.1800	4.7372
BZM16	−3.9000	308.1380	1.3200	5.5400	4.0340
BZM19	−4.2359	293.4540	1.3035	5.5200	3.5960
BZM20	−3.6134	204.8540	1.3147	5.1900	5.5445
BZM21	−3.6102	214.6073	1.3143	5.6200	5.7093
			Test Set		
BZM6	−3.6587	256.3247	1.3018	4.2000	4.1185
BZM7	−3.8331	283.6173	1.3017	5.4000	4.5022
BZM11	−4.1378	338.7730	1.3020	7.1600	4.2642
BZM14	−4.1352	322.5207	1.3102	5.6300	3.8790
BZM17	−3.7969	230.3647	1.3142	5.0500	4.6922
BZM18	−3.9513	239.7053	1.3027	3.7600	3.8227
BZM22	−3.1385	218.2873	1.3140	5.6200	6.5058

The equation of the QSAR model is presented below. The statistical indicators are given in Table 14.

$$\begin{aligned}\text{pMIC}^{\text{th}} &= 5.18377051213999 - 7.94306843387908 * X_1 \\ &+9.74799637991962 * X_2 + 6.75806419946355 * X_3\end{aligned}$$

With:

$$\begin{aligned}X_1 = \text{TanH}\big(&0.5\big((-17.676793613319) - 0.585007547192468 * \mu \\ &+ 0.0129725702462401 * \alpha + 5.65823485396223 * l\big(\text{C=N}\big) \\ &+ 0.84970063740487 * \log P\big)\end{aligned}$$

$$\begin{aligned}X_2 = \text{TanH}\big(&0.5 * \big((-45.1752463091173) - 1.24094592928177 * \mu \\ &- 0.00456140471899028 * \alpha + 30.772101165482 * l\big(\text{C=N}\big) \\ &+ 0.214326830401005 * \log P\big)\end{aligned}$$

$$\begin{aligned}X_3 = \text{TanH}\big(&0.5 * \big(52.1085892484661 + 1.76468372023022 * \mu \\ &+ 0.0184209118191153 * \alpha - 41.7842796272828 * l\big(\text{C=N}\big) \\ &+ 0.846662473397522 * \log P\big)\end{aligned}$$

The coefficient of determination (R^2 = 0.9995), shows that the predicted pMIC values contain 96.48% of the experimental values. The value of the Fisher test (F = 31879.0548) is very high compared to the critical value, from the Fisher table Fcr = 2.96 [50]. This value 31879.0548 of Fisher's test, higher than the critical value, shows that the error committed is lower than what the model explains [50]. The standard deviation (RMSE = 0.0149) expresses the small deviation of the predicted values from the experimental mean. This model has a cross-validation correlation coefficient Q^2_{cv} equal to $Q^2_{cv} = 0.9995$. This value, higher than 0.9, reflects a so-called excellent model according to Erikson *et al.* [34]. Ce modèle est acceptable car il is in agreement with the acceptance criteria of these authors $R^2 - Q^2_{cv} = 0.9995 - 0.9995 = 0.000 < 0.3$. All these statistical indicators show that the model developed explains the TB activity in a statistically significant and satisfactory manner. These different results are confirmed by the regression plot of the ANN model presenting the theoretical anti-tuberculosis activity as a function of the experimental activity represented in Figure 10.

Table 14. Statistical analysis ratio of potential inhibitory concentration (pMIC) of benzimidazole derivatives of ANN model.

Number of observations N	15
Coefficient of determination R^2	0.9995
Standard deviation RMSE	0.0149
Fischer test F	31879.0548
Cross-validation correlation coefficient Q^2_{CV}	0.9995
Confidence level α	>95%

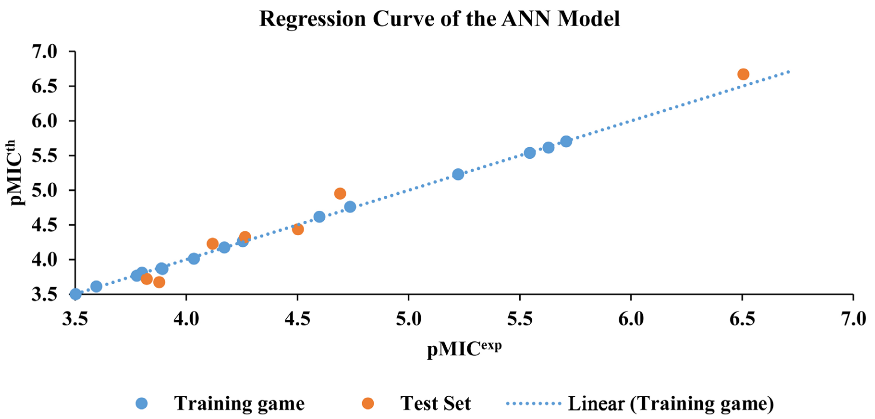

Figure 10. The regression line of the ANN model.

The regression curve of the ANN model shows that all points are around the regression line. This result indicates that there is a small difference (RMSE = 0.0149) between the values of pMIC^{exp} and pMICth, thus a good similarity in these values. This similarity is illustrated in Figure 11.

1) Internal Validation

a) Leave-One-Out (LOO) Procedure

The internal validation of the RML model was performed using the leave-one-out (LOO) cross validation technique on the 15 molecules of the training set. The results obtained are gathered in Table 15.

The results show that the models constructed, after the removal of one of the compounds from the training set (first column of the table), have statistical parameters (R^2 and RMSE) of the same order as those of the initial model, overall. The average values of these parameters are R^2_{LOO} = **0.9661**, RMSE$_{\text{LOO}}$ = **0.1216**. We find values almost identical to those of the initial model. The cross-correlation coefficient $Q^2_{cv\,\text{LOO}}$, is equal to 0.7936. This value is higher than the minimum required value of 0.5 according to Tropsha *et al.* [51] [52]. In addition, we note that $R^2_{\text{LOO}} - Q^2_{cv\,\text{LOO}} = 0.1283 < 0.3$ [53]. All this shows that the ANN model has a very satisfactory internal predictive character and can be considered as robust [54].

b) Randomization Test

The randomization test of the ANN model was performed on the molecules of the training set by randomly permuting the values of the activities while keeping the descriptors for model building. We stopped at ten (10) iterations. The randomized coefficients of determination (R^2_r) for each iteration are listed in Table 16.

From the values in Table 16, the value of Roy's parameter ($R^2_p = 0.5445$) was determined. This value ($R^2_p = 0.5445$) is lower than the coefficient of determination of the model (0.9995). These different results show that the is not due to chance and can be considered as robust.

2) External Validation

External validation of the ANN model was performed on the molecules in the

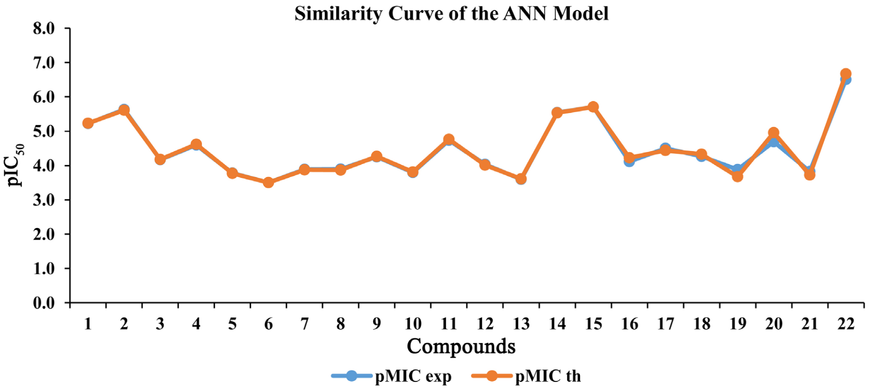

Figure 11. Similarity curve of experimental and predicted values of the ANN model.

Table 15. Statistical parameters of the leave-one-out (LOO) cross-validation of the ANN model.

MOLECULES	pMIC^{exp}	pMIC^{pred}	R^2	RMSE	F	Q^2_{LOO}
BZM1	5.2224	5.2343	0.9404	0.1797	189.5224	**0.9609**
BZM2	5.629	5.3197	0.9467	0.1599	240.4986	
BZM3	4.171	4.0152	0.9438	0.1815	233.5587	
BZM4	4.5991	4.5930	0.9551	0.1625	260.2607	
BZM5	3.7777	3.9481	0.9622	0.1454	305.6039	
BZM8	3.5029	3.6221	0.9738	0.1174	446.1816	
BZM9	3.889	3.8686	0.9546	0.1607	252.5907	
BZM10	3.8927	3.8648	0.9647	0.1419	327.7217	
BZM12	4.2546	4.1230	1	1.19E−09	5.4754E+18	
BZM13	3.8012	3.6863	0.9999	0.0032	643424.4387	
BZM15	4.7372	4.9377	0.9552	0.1617	256.0626	
BZM16	4.034	4.0750	0.9671	0.138	353.3991	
BZM19	3.596	3.8548	0.9438	0.174	204.1541	
BZM20	5.5445	5.6447	0.9996	0.0147	31774.1965	
BZM21	5.7093	5.6137	0.9852	0.0828	801.8245	
Averages			**0.9661**	**0.1216**	**3.65E+17**	

Table 16. Randomized coefficients of determination (R_r^2) of the ten (10) iterations.

Iteration	1	2	3	4	5	6	7	8	9	10
R_r^2	0.8082	0.7445	0.8252	0.7422	0.6354	0.6975	0.7506	0.5905	0.6283	0.6042

validation set (**Table 13**) using the Tropsha criteria [51] [52] and Roy [38]. The Tropsha and Roy criteria checks are recorded in **Table 17** and **Table 18** respectively.

The values in **Table 17** show that all Tropsha criteria are met, so the model is acceptable for predicting the antitubercular activity of benzimidazole derivatives.

The analysis in Table 18 shows that r_m^2 is greater than 0.5 and the Δr_m^2 is less than 0.2. This result reflects that the model meets Roy's criteria. We can therefore affirm that the model is robust and has good predictive power.

3.2.6. Domain of Applicability of the ANN Model

The applicability domain of the model was determined by the leverage method. The hii-lever values of the molecules in the training set calculated from the MINTAB software are listed in Table 19.

The values of the h_{ii} levers in Table 19 and the standardized residues of the molecules were used to plot the graph of standardized residues versus h_{ii} levers in Figure 12.

Table 17. Tropsha criterion checks of the ANN model external test set.

Statistical parameters	Tropsha criteria [35] [36] [37]	
R^2	>0.7	**0.9827**
Q_{CV}^2	>0.6	**0.9671**
$\left\|R^2 - R_0^2\right\|$	≤0.3	**0.007**
$\frac{\left\|R^2 - R_0^2\right\|}{R^2}$	<0.1	**0.007**
k	$0.85 \le k \le 1.15$	**1.0105**
$\frac{\left\|R^2 - R_0'^2\right\|}{R^2}$	<0.1	**0.012**
k'	$0.85 \le k' \le 1.15$	**0.9886**

Table 18. Roy criteria checks of the external validation set of the ANN model.

Indicators	r_m^2	$r_m'^2$	$\overline{r_m^2} = \frac{r_m^2 + r_m'^2}{2}$	$\Delta r_m^2 = \left\|r_m^2 - r_m'^2\right\|$
Value	**0.902**	**0.876**	**0.889**	**0.026**

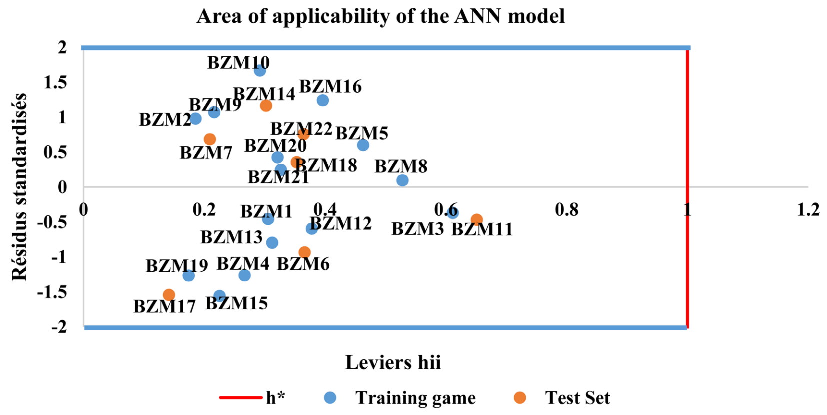

Figure 12. Graph of standardized residuals of antituberculosis activity according to the levers of the ANN model.

Table 19. Lever values of the molecules in the training set.

Molecules	h_{ii}
BZM1	0.3055
BZM2	0.1855
BZM3	0.6113
BZM4	0.2664
BZM5	0.4629
BZM8	0.5277
BZM9	0.2162
BZM10	0.2921
BZM12	0.3776
BZM13	0.3121
BZM15	0.2250
BZM16	0.3961
BZM19	0.1739
BZM20	0.3212
BZM21	0.3267

For the 15 molecules of the training set and the 4 descriptors of the model, the threshold value of the h^* levers is 1. The extreme values of the standardized residuals are ±3 according to the "three sigma rule" [49]. These different values delimit the field of applicability [55] of the model as shown on the graph in **Figure 12**. **Figure 12** shows us that all the molecules have levers lower than the threshold lever ($h^* = 1$) and standardized residue values between +3 and −3. This result means that all the molecules belong to the applicability domain.

4. Conclusion

At the end of this work we determined three (3) mathematical relationships between the potential minimum inhibitory concentration (pMIC) of benzimidazole derivatives against *Mycobacterium tuberculosis strain H*37*Rv* and their physicochemical descriptors. Chemical potential (μ), polarizability (α), bond length l (C = N), and lipophilicity (log$P$) are the parameters that explain the antitubercular activity of benzimidazole derivatives significantly. Statistical data learning methods such as multilinear regression (MLR), nonlinear multiple regression (NLMR) as well as artificial neural network (ANN) methods were used. The statistical indicators of these models (MLR, NLMR, ANN) show that they are acceptable, robust and have good predictive power. In this work, due to its statistical indicators, the artificial neuron method (ANN) ($R^2 = 0.9995$; RMSE = 0.0149; $F = 31879.0548$) proved to be the best statistical learning method for predicting the anti-tuberculosis activity of benzimidazole derivatives. Moreover, the applicability range of this model determined from the levers shows that a prediction of the pMIC of the new benzimidazole derivatives is acceptable when its lever value is less than 1. In perspective we intend to use the artificial neural network model to predict the biological activity of new benzimidazole derivatives.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Institut Pasteur. https://www.pasteur.fr/fr/centre-medical/fiches-maladies/tuberculose

[2] OMS (Organisation Mondiale de la Santé) (2021) Tuberculose. https://www.who.int/fr/news-room/fact-sheets/detail/tuberculosis

[3] Gabo, G. (2021) Prévalence de la tuberculose en Côte d'Ivoire: 19 976 porteurs de la maladie dépistés en 2020. https://www.fratmat.info/article/211803/societe/santeacute/prevalence-de-la-tuberculose-en-cote-divoire-19-976-porteurs-de-la-maladie-depistes-en-2020

[4] Castagnolo, D., Manetti, F., Radi, M., Bechi, B., Pagano, M., De Logu, A., Meleddu, R., Saddi, M. and Botta, M. (2009) Synthesis, Biological Evaluation, and SAR Study of Novel Pyrazole Analogues as Inhibitors of *Mycobacterium tuberculosis*: Part 2. Synthesis of Rigid Pyrazolones. *Bioorganic & Medicinal Chemistry*, **17**, 5716-5721. https://doi.org/10.1016/j.bmc.2009.05.058

[5] Raynaud, C., Daher, W., Johansen, M., Roquet-Baneres, F., Blaise, M., Onajole, O., Kozikowski, A., Herrmann, J.-L., Dziadek, J., Gobis, K. and Kremer, L. (2020) Active Benzimidazole Derivatives Targeting the MmpL3 Transporter in *Mycobacterium abscessus*. *ACS Infectious Diseases*, **6**, 324-337. https://doi.org/10.1021/acsinfecdis.9b00389

[6] Garuti, L., Roberti, M., Malagoli, M., Rossi, T. and Castelli, M. (2001) Synthèse et activité antiproliférative de certains thiazolylbenzimidazole-4,7-diones. *Bioorganic & Medicinal Chemistry Letters*, **11**, 3147-3149. https://doi.org/10.1016/S0960-894X(01)00639-4

[7] Rao, A., Chimirri, A., De Clercq, E., Monforte, A., Monforte, P., Pannecouque, C. and Zappala, M. (2002) Synthesis and Anti-HIV Activity of 1-(2,6-difluorophenyl)-1H,3H-thiazolo[3,4-a]benzimidazole Structurally-Related 1,2-substituted Benzimidazoles. *Il Farmaco*, **57**, 819-823. https://doi.org/10.1016/S0014-827X(02)01300-9

[8] Valdez, J., Cedillo, R., Hernández-Campos, A., Yépez, L., Hernández-Luis, F., Navarrete-Vázquez, G., Tapia, A., Cortés, R., Hernández, M. and Castillo, R. (2002) Synthesis and Antiparasitic Activity of 1*H*-Benzimidazole Derivatives. *Bioorganic & Medicinal Chemistry Letters*, **12**, 2221-2224. https://doi.org/10.1016/S0960-894X(02)00346-3

[9] Thakurdesai, P., Wadodkar, S. and Chopade, C. (2007) Synthesis and Anti-Inflammatory Activity of Some Benzimidazole-2-Carboxylic Acids. *Pharmacologyonline*, **1**, 314-329.

[10] Ayhan-Kilcigil, G., Kus, C., Ozdamar, E., Can-Eke, B. and Iscan, M. (2007) Synthesis and Antioxidant Capacities of Some New Benzimidazole Derivatives. *Archiv der Pharmazie*, **340**, 607-611. https://doi.org/10.1002/ardp.200700088

[11] Serafin, B., Borkowska, G., Główczyk, J., Kowalska, I. and Rump, S. (1989) Potential Antihypertensive Benzimidazole Derivatives. *Polish Journal of Pharmacology and Pharmacy*, **41**, 89-96.

[12] Oprea, T.I. (2005) Chemoinformatics in Drug Discovery. Wiley-VCH Verlag GmbH & Co. KGaA, Hoboken. https://doi.org/10.1002/3527603743

[13] Rekka, E.A. and Kourounakis, P.N. (2008) Chemistry and Molecular Aspects of Drug Design and Action. CRC Press, Boca Raton. https://doi.org/10.1201/9781420008272

[14] Frisch, M.J., Trucks, G.W., Schlegel, H.B. and Scuseria, G.E. (2009) Gaussian 09, Revision A.02. Gaussian, Inc., Wallingford.

[15] Chattaraj, P.K., Cedillo, A. and Parr, R.G. (1991) Variational Method for Determining the Fukui Function and Chemical Hardness of an Electronic System. *The Journal of Chemical Physics*, **103**, 7645-7646. https://doi.org/10.1063/1.470284

[16] Ayers, P.W. and Parr, R.G. (2000) Variational Principles for Describing Chemical Reactions: The Fukui Function and Chemical Hardness Revisited. *Journal of the American Chemical Society*, **122**, 2010-2018. https://doi.org/10.1021/ja9924039

[17] De Proft, F.J., Martin, M.L. and Geerlings, P. (1996) On the Performance of Density Functional Methods for Describing Atomic Populations, Dipole Moments and Infrared Intensities. *Chemical Physics Letters*, **250**, 393. https://doi.org/10.12980/APJTB.4.2014C1012

[18] Hansch, C., Sammes, P.G. and Taylor, J.B. (1990) Comprehensive Medicinal Chemistry. *Computers and the Medicinal Chemist*, Vol. 4, Pergamon Press, Oxford, 33-58.

[19] Franke, R. (1984) Theoretical Drug Design Methods. Elsevier, Amsterdam.

[20] Chaltterjee, S., Hadi, A. and Price, B. (2000) Regression Analysis by Examples. Wiley VCH, New York.

[21] Phuong, H. (2007) Synthèse et étude des relations structure/activité quantitatives (QSAR/2D) d'analoguesBenzo[c]phénanthridiniques. France.

[22] Microsoft Excel, Miicrosoft Office Version 2016.

[23] XLSTAT Version (2014) XLSTAT and Addinsoft are Registered Trademarks of Addinsoft.

[24] JMPPro13 (2014) Statistical Discovery. SAS Institute Inc., Scintilla, 1998-2014.

[25] Tammo (1995) Theoretical Analysis of Molecular Membrane Organization. CRC, Raton.

[26] Rutkowska, E., Pajak, K. and Jozwiak, K. (2013) Lipophilicity—Methods of Determination and Its Role in Medicinal Chemistry. *Acta Poloniae Pharmaceutica*: *Drug Research*, **70**, 3-18.

[27] Bakht, M.A., Alajmi, M.F., Alam, P., Alam, A., Alam, P. and Aljarba, T.M. (2014) Theoretical and Experimental Study on Lipophilicity and Wound Healing Activity of Ginger Compounds. *Asian Pacific Journal of Tropical Biomedicine*, **4**, 329-333. https://doi.org/10.12980/APJTB.4.2014C1012

[28] Kujawski, J., Popielarska, H., Myka, A., Drabińska, B. and Bernard, M.K. (2012) The logP Parameter as a Molecular Descriptor in the Computer-Aided Drug Design—An Overview. *Computational Methods in Science and Technology*, **18**, 81-88. https://doi.org/10.12921/cmst.2012.18.02.81-88

[29] Cozma, A., Zaharia, V., Ignat, A., Gocan, S. and Grinberg, N. (2012) Prediction of the Lipophilicity of Nine New Synthesized Selenazoly and Three Aroyl-Hydrazinoselenazoles Derivatives by Reversed-Phase High Performance Thin-Layer Chromatography. *Journal of Chromatographic Science*, **50**, 157-161.

https://doi.org/10.1093/chromsci/bmr034

[30] Acdlabs (2010) Advanced Chemistry Development/Chemskecht, 1994-2010.

[31] Snedecor, G.W. and Cochran, W.G. (1967) Methods, Statistical. Oxford and IBH, New Delhi, 381.

[32] Kangah, N.J.-B., Koné, M.G.-R., Kodjo, C.G., N'guessan, B.R., Kablan, A.L.C., Yéo, S.A. and Ziao, N. (2017) Antibacterial Activity of Schiff Bases Derived from Ortho Diaminocyclohexane, Meta-Phenylenediamine and 1,6-Diaminohexane: Qsar Study with Quantum Descriptors. *International Journal of Pharmaceutical Science Invention*, **6**, 38-43.

[33] Esposito, E.X., Hopfinger, A.J. and Madura, J.D. (2004) Methods for Applying the Quantitative Structure-Activity Relationship Paradigm. In: Bajorath, J., Ed., *Chemoinformatics*, Vol. 275, Humana Press, Totowa, 131-213. https://doi.org/10.1385/1-59259-802-1:131

[34] Eriksson, L., Jaworska, J., Worth, A., Cronin, M.D., Mc Dowell, R.M. and Gramatica, P. (2003) Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification- and Regression-Based QSARs. *Environmental Health Perspectives*, **111**, 1361-1375. https://doi.org/10.1289/ehp.5758

[35] Golbraikh, A. and Tropsha, A. (2002) Beware of q^2. *Journal of Molecular Graphics and Modelling*, **20**, 269-276. https://doi.org/10.1016/S1093-3263(01)00123-1

[36] Tropsha, A., Gramatica, P. and Gombar, V.K. (2003) The Importance of Being Earnest, Validation Is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR & Combinatorial Science*, **22**, 69-77. https://doi.org/10.1002/qsar.200390007

[37] Ouattara, O., Affi, T.S., Koné, M.G.-R., Bamba, K. and Ziao, N. (2017) Can Empirical Descriptors Reliably Predict Molecular Lipophilicity? A QSPR Study Investigation. *International Journal of Engineering Research and Application*, **7**, 50-56. https://doi.org/10.9790/9622-0705015056

[38] Roy, P.P. and Roy, K. (2008) On Some Aspects of Variable Selection for Partial Least Squares Regression Models. *QSAR & Combinatorial Science*, **27**, 302-313. https://doi.org/10.1002/qsar.200710043

[39] Hea, G., Fenga, L. and Chena, H. (2012) International Symposium on Safety Science and Engineering in China. *Procedia Engineering*, **43**, 204-209.

[40] Dreyfus, G. (1998) Réseaux de neurones artificiels. Toulouse.

[41] Dreyfus, G., Martinez, J., Samuelides, M., Gordon, M., Badran, F., Thiria, S.and Herault, L. (2002) Réseaux de Neurones Artificiels. 2 édition, Groupe Eyrolles, New York, 374. https://doi.org/10.1177/026119290503300510

[42] Jeliazkova, N.N. and Jaworska, J. (2005) An Approach to Determining Applicability Domains for QSAR Group Contribution Models: An Analysis of SRC KOWWIN. *Alternatives to Laboratory Animals*, **33**, 461-470. https://doi.org/10.1177/026119290503300510

[43] Sahigara, F., Mansouri, K., Ballabio, D., Mauri, A. and Todeschini, V.C.a.R. (2012) Comparison of Different Approaches to Define the Applicability Domain of QSAR Models. *Molecules*, **17**, 4791-4810. https://doi.org/10.3390/molecules17054791

[44] Roy, K. (2015) Chapter 2. Statistical Methods in QSAR/QSPR. In: *A Primer on QSAR/QSPR Modeling*, Springer, Cham, 37-59. https://doi.org/10.1007/978-3-319-17281-1_2

[45] Jaworska, J., Jeliazkova, N.N. and Aldenberg, T. (2005) QSAR Applicability Domain Estimation by Projection of the Training Set in Descriptor Space: A Review. *Alter-*

natives to Laboratory Animals, **33**, 445-459. https://doi.org/10.1155/2016/5137289

[46] Ghamali, M., Chtita, S., Bouachrine, M. and Lakhlifi, T. (2016) Méthodologie générale d'une étude RQSA/RQSP. *Revue Interdisciplinaire*, **1**, 1-6.

[47] Chtita, S., Ghamali, M., Hmamouchi, R., Elidrissi, B., Bourass, M., Larif, M., Bouachrine, M. and Lakhlifi, T. (2016) Investigation of Antileishmanial Activities of Acridines Derivatives against Promastigotes and Amastigotes form of Parasites Using QSAR Analysis. *Advances in Physical Chemistry*, **2016**, Article ID: 5137289. https://doi.org/10.1155/2016/5137289

[48] Asadollahi, T., Dadfarnia, S., Shabani, A., Ghasemi, J. and Sarkhosh, M. (2011) QSAR Models for CXCR2 Receptor Antagonists Based on the Genetic Algorithm for Data Preprocessing Prior to Application of the PLS Linear Regression Method and Design of the New Compounds Using *in Silico* Virtual Screening. *Molecules*, **16**, 1928-1955. https://doi.org/10.3390/molecules16031928

[49] Chtita, S., Larif, M., Ghamali, M., Bouachrine, M. and Lakhlifi, T. (2015) Quantitative Structure-Activity Relationship Studies of dibenzo[*a*,*d*]cycloalkenimine Derivatives for Non-Competitive Antagonists of *N*-Methyl-D-Aspartate Based on Density Functional Theory with Electronic and Topological Descriptors. *Journal of Taibah University for Science*, **9**, 143-154. https://doi.org/10.1016/j.jtusci.2014.10.006

[50] Fortuné, A. (2006) Techniques de Modélisation Moléculaire appliquées à l'Etude et à l'Optimisation de Molécules Immunogènes et de Modulateurs de la Chimiorésistance. Médicaments.

[51] Tropsha, A. (2010) Best Practices for QSAR Model Development, Validation, and Exploitation. *Molecular Informatics*, **29**, 476-488. https://doi.org/10.1002/minf.201000061

[52] Shamsara, J. (2017) Ezqsar: An R Package for Developing QSAR Models Directly From Structures. *The Open Medicinal Chemistry Journal*, **11**, 212-221. https://doi.org/10.1590/S0103-50532009000400021

[53] Kiralj, R., Ferreira, M.M.C. and Braz, J. (2009) Basic Validation Procedures for Regression Models in QSAR and QSPR Studies: Theory and Application. *Journal of the Brazilian Chemical Society*, **20**, 770-787. https://doi.org/10.1590/S0103-50532009000400021

[54] Veerasamy, R., Rajak, H., Jain, A., Sivadasan, S., Varghese, C.P. and Agrawal, R.K. (2011) Validation of QSAR Models—Strategies and Importance. *International Journal of Drug Design and Discovery*, **2**, 511-519.

[55] Gramatica, P. (2007) Principles of QSAR Models Validation: Internal and External. *QSAR & Combinatorial Science*, **26**, 694-701. https://doi.org/10.1002/qsar.200610151