



# IN SILICO STUDY FOR INVESTIGATING AND PREDICTING THE ACTIVITIES OF 1,2,4-TRIAZOLE DERIVATIVES AS POTENT ANTI-TUBERCULAR AGENTS

A. E. SHOLA<sup>1,\*</sup>, S. UBA<sup>1</sup> and A. UZAIRU<sup>1</sup>

<sup>1</sup>Ahmadu Bello University, Department of Chemistry, Zaria, Kaduna State, Nigeria

\*Corresponding author. Ahmadu Bello University, Department of Chemistry, Zaria, Kaduna State, Nigeria, Phone: +2347060511720  
e-mail address: [shola4343@gmail.com](mailto:shola4343@gmail.com) (A. E. SHOLA).

## ARTICLE INFO

Article history:

Received 2018-01-03

Accepted 2018-01-31

Available online 2018-06-30

### keywords

Tuberculosis

1,2,4-Triazole

QSAR

Applicability domain

Y-Randomization

## ABSTRACT

*In silico* study was carried on a dataset of 1,2,4-Triazole derivatives to investigate their activities behaviour on mycobacterium tuberculosis by utilizing Quantitative Structure-Activity Relationship (QSAR) technique. Genetic Function Algorithm (GFA) and Multiple Linear Regression Analysis (MLRA) were used to select the optimum descriptors and to generate the correlation QSAR model that relate their activities values against mycobacterium tuberculosis with the molecular structures of the inhibitors. The model was validated and was found to have squared correlation coefficient ( $R^2$ ) of 0.9134, adjusted squared correlation coefficient ( $R_{adj}$ ) of 0.8753 and Leave one out (LOO) cross validation coefficient ( $Q_{cv}^2$ ) value of 0.8231. The external validation set used for confirming the predictive power of the model has  $R^2_{pred}$  of 0.7482. Stability and robustness of the model obtained by the validation test indicate that the model can be used to design and synthesis other 1,2,4-Triazole derivatives with improved anti-mycobacterium tuberculosis activities.

## 1. INTRODUCTION

Tuberculosis (TB) is infectious disease caused by *Mycobacterium tuberculosis*. About 2.5 billion people were infected with tuberculosis worldwide and mortality of approximately 1.5 million people were reported annually (Jham et al., 2014; Lönnroth et al., 2010). Despite the availability of tuberculosis first-line drug; ethambutol (EMB), rifampicin (RIF), pyrazinamide (PZA), streptomycin (STP) and isoniazid (INH); the increase in the incidence of both multidrug-resistant (MDR-TB) and extensively drug-resistant tuberculosis (XDR-TB) are observed (Aziz et al., 2006; Balabanova et al., 2005). Furthermore, treatment requiring the use of these drugs has been reported to cause serious side effects such as: thrombocytopenia occurring as a result of rifampicin (RIF) (Yakar et al., 2013), neuropathy is caused by isoniazid while biggest problem associated with hepatitis (Abideen et al., 2013). Considering these effect, the synthesis of new compounds with anti-tuberculosis activity has been the target of many medicinal chemistry and pharmacist.

Recently, a novel series of 1,2,4-Triazole derivatives has been identified and reported as inhibitors against of *M. tuberculosis* (Sarkar et al., 2016). The advent of computational chemistry and knowledge of theoretical chemistry led to challenges of drug discovery (Cramer, Patterson and Bunce, 1988).

QSAR as computational tool establish a relationship between various molecular properties of molecules and their activities (Ibezim et al., 2009). The application of Quantitative Structure Activity Relationship (QSAR) technique has potential to minimize effort and time required to discover new compounds or to improve current ones in terms of their efficiency.

The aim of this research was to develop QSAR model using Genetic Function Algorithm (GFA) for variable selection of descriptors and multiple linear regression (MLR) method for predicting the activity of 1,2,4-Triazole derivatives as potent anti-mycobacterium tuberculosis.

## 2. MATERIALS AND METHOD

### 2.1 Data set

The derivatives of 1,2,4-Triazole derivatives as potent anti-mycobacterium tuberculosis that were used in this research were selected from the literature (Sarkar et al., 2016). The anti-mycobacterium tuberculosis were expressed in percentage (%) and later converted to logarithm scale using the Equation (1) below in order to increase the linear response, reduce the dispersion and approach normal distribution. The general formula and chemical structures alongside with their biological activities of these compounds were presented Table A1.

$$pBA = \log \left[ \left( \frac{\text{Molecular weight} \left( \frac{g}{mol} \right)}{\text{Dose} \left( \frac{g}{mol} \right)} \right) \left( \frac{\text{percentage} (\%)}{100 - \text{percentage} (\%)} \right) \right] \quad (1)$$

### 2.2 Molecular structure optimization and descriptors calculation

Molecular structures of the dataset compounds were optimized using the Density Function Theory (DFT) with basic (B3LYP/631G\*) by employing Spartan 14. The optimized

structures were imported into PaDEL-Descriptors software to calculate about 1875 molecular descriptors. The calculated descriptors and the activity values of the compounds were arranged in  $n \times m$  matrix format to constitute the data used for the QSAR study. Where  $n$  is the number of molecules and  $m$  is the number of descriptors.

### 2.3 Descriptors transformation

In QSAR analysis, model is usually biased toward descriptor with high positive or negative values. To overcome this bias, molecular descriptors of the training set data were transformed by normalization (Tropsha, 2010) using the equation below;

$$x^n = \frac{x - x_{max}}{x_{max} - x_{min}} \quad (2)$$

In equation 2,  $x^n$  is the normalized descriptor,  $x_{max}$  is the maximum value in a descriptor column and  $x_{min}$  is the minimum value in the column.

### 2.4 Data pretreatment

The data set were subjected to pretreatment using Data Pretreatment software obtained from Drug Theoretical and Cheminformatics Laboratory (DTC Lab) in order to remove redundant data. In the dataset matrix, all descriptors column with constant or near constant values were deleted. Also, anyone column with variance less than 0.001 was deleted from the descriptor pool. Correlation analysis was performed on the dataset and one descriptor of any pair with a correlation greater than 0.8 was discarded.

### 2.5 Generation of training set and test set

In order to obtain validated QSAR model, the dataset was divided into training and test set using Data Division software obtained from Drug Theoretical and Cheminformatics Laboratory (DTC Lab) by employing Kennard and Stone's algorithm (Kennard and Stone, 1969). This algorithm has been applied with great success in many recent QSAR studies and has been highlighted as one of the best ways to build training and test set (Afantitis *et al.*, 2006; Chakraborti *et al.*, 2003; Khaled, 2011; Melagraki *et al.*, 2006; Wu *et al.*, 1996). In this algorithm, two compounds with the largest Euclidean distance apart were initially selected for the training set. The remaining compounds for the training set were selected by maximizing the minimum distance between these two compounds and the rest of the compounds in the dataset. This process continues until the desired number of compounds needed for the training set have been selected then, the remaining compounds in the dataset would be used as the test set (Kennard and Stone, 1969). The algorithm employs Euclidean distance  $ED_x(p, q)$ , between the  $x$  vectors of each pair  $(p, q)$  of samples to ensure a uniform distribution of such a subset along the  $x$  data space.

$$ED_x(p, q) = \sqrt{\sum_{j=1}^N [x_p(j) - x_q(j)]^2}, p, q \in [1, M] \quad (3)$$

$N$  is the number variables in  $x$ , and  $M$  is the number of samples while  $x_p(j)$  and  $x_q(j)$  are the  $j$ th variable for samples  $p$  and  $q$  respectively. The training set was used to generate the model, while the test set were used for the external validation of the model

## 2.6 Selection of optimal descriptors

Genetic Function Algorithm (GFA) incorporated in Material Studio software version 8.0 was used to select combination of descriptors that best correlate the structure of the inhibitory compounds with their respective activities. GFA is a heuristic search method that finds exact or approximate solutions to any optimization and search problems (Rogers and Hopfinger, 1994). It has the advantage of producing more than one combination of descriptors using lack of fit function to eliminate over fitting and gave user control over length of equation.

## 2.7 Model construction and validation

The optimum descriptors combinations selected by GFA were obtained for both training set from the descriptor pool. Their anti-tubercular activities were placed as the last column in their respective spread sheets which were imported into the in Material Studio software version 8.0 to generate the model and to evaluate the internal validation parameters while the test set was used to validate the built model.

## 2.8 Relative importance of each descriptor in the model

Absolute value of the mean effect of each descriptor was used to evaluate the relative importance and contribution of the descriptor to the model. The mean effect is defined as;

$$ME = \frac{\beta_j \sum_i^n D_j}{\sum_j^m (\beta_j \sum_i^n D_j)} \quad (4)$$

Where ME is the mean effect of a descriptor  $j$  in a model,  $\beta_j$  is the coefficient of the descriptor  $j$  in that model,  $D_j$  is the value of the each descriptor in the data matrix for each molecule in the training set and  $m$  is the number of descriptor that appear in the model and  $n$  is the number of molecules in the training set (Minovski *et al.*, 2013)

## 2.9 Models applicability domain

Leverage approach that utilizes Williams plot was employed to define the applicability domain of the models and to determine the influential molecule (Tropsha *et al.*, 2003). Williams plot for a model is a graphical view of leverage values for each molecule in the entire data set versus their standardized cross validated residual obtained by the model. The leverage ( $h_i$ ) value for each molecule is obtained has the diagonal elements of the hat matrix constructed for both training set and test set. Leverage of a given chemical compound  $h_i$ , is defined as follows;

$$h_i = X_i(X^T X)^{-1} X_i^T \quad (5)$$

Where  $X_i$  is training compounds matrix of  $i$ .  $X$  is the  $m \times k$  descriptor matrix of the training set compound and  $X^T$  is the transpose matrix of  $X$  used to build the model. The warning leverage ( $h^*$ ) is the boundary of values for  $X$  outliers and is defined as:

$$h^* = 3 \frac{(d+1)}{m} \quad (6)$$

Where  $m$  is the descriptors and  $d$  is the compound that made up the training set.

## 2.10 Y-Randomization test

Y-Randomization test is another useful external validation parameter to confirm that the built QSAR model is strong and not inferred by chance. To be assured that the built QSAR model is strong, reliable and not obtained by chance, the Y-randomization test was carried out on the compound that made up the training set (Tropsha *et al.*, 2003). For the built QSAR model to robust and reliable, the model is expected to have a low  $R^2$  and  $Q^2$  values for several trials. Coefficient of determination ( $cR_p^2$ ) for Y-randomization is another parameter calculated which should be greater than 0.5 for passing this test.

$$cR_p^2 = R \times [R^2 - (R_r)^2]^2 \quad (7)$$

$cR_p^2$  is Coefficient of determination for Y-randomization,  $R$  is coefficient of determination for Y-randomization and  $R_r$  is average 'R' of random models.

## 3. RESULTS AND DISCUSSION

*In silico* study was carried to develop a QSAR model that could relate the structure of 1, 2, 4-Triazole derivatives with their biological activities against M. tuberculosis.

Experimental and Predicted activities of 1, 2, 4-Triazole derivatives and the residual values were presented in Table 1. The low residual value between Experimental and Predicted activities indicates that the model has a good predictive power.

The Genetic Function Algorithm (GFA) method employed in this study led to the selection of three descriptors which were used to build a model for calculating the predicted activities against M. tuberculosis.

### 3.1 QSAR model generated

The QSAR model generated by the combinations of methods used in the study is presented below:

$$pBA = -0.314545676 (\text{AATS7s}) + 0.016884210 (\text{TDB9e}) + 0.09032096 (\text{RDF90i}) + 4.023544336 \quad (8)$$

All the validation parameters to confirm the stability, robustness and predictive ability of the model generated were reported in in Table 1.

The name and symbol of the three descriptors used in generating the QSAR model were reported in Table 2. The presence of 3D and 2D descriptors in the model suggest that these types of descriptors were able to characterize better anti-mycobacterium tuberculosis activities.

Pearson's correlation and statistics of the three descriptors used in the QSAR model were reported in Table 3 which shows clearly that the correlation coefficients between each descriptor is very low thus, it can be inferred that there is no significant inter-correlation among the descriptors used in building the model. The Variance Inflation Factor (VIF) value for each the descriptor in the model was less than 4 which imply that the Model was statistically significant and the descriptors were orthogonal. The signs and the magnitude of these descriptors combined with their mean effects indicate their individual strength and direction in influencing the activity of a compound. The null hypothesis says there is no significant relationship between the activities of the inhibitor molecules and the descriptors used in building the model at  $p > 0.05$ . The P-values of the descriptors in the model at 95% confidence limit

shown in Table 3 are all less than 0.05. This implies that the null hypothesis is rejected. Thus we accepted the alternative hypothesis. Hence we infer that there is a significant relationship between the activities of the inhibitor molecules and descriptors used in building the model at  $p < 0.05$ .

Y- Randomization parameter test were reported in Table 4. The low  $R^2$  and  $Q^2$  values for several trials assured that the developed QSAR model is robust, stable and reliable. While the  $cR_p^2$  value greater than 0.5 satisfied that the built model is powerful and not inferred by chance.

**Table 1 – Internal and External validation parameters for the QASR model generated**

Parameter	Formula	Threshold	Model score	comment	Reference
<b>Internal validation</b>					
$R^2$	$\frac{[\sum \{(Y - \bar{Y}) \times (\hat{Y} - \bar{Y})\}]^2}{\sum (Y - \bar{Y})^2 \times \sum (\hat{Y} - \bar{Y})^2}$	$R^2 > 0.6$	0.9134	passed	(Tropsha, 2010)
$R_{adj}^2$	$\frac{(N - 1) \times R^2 - p}{N - 1 - p}$	$R_{adj}^2 > 0.6$	0.8753	passed	
$Q^2$	$1 - \frac{\sum (Y - \hat{Y}_{100})^2}{\sum (Y - \bar{Y})^2}$	$Q^2 > 0.6$	0.8231	Passed	
$F_{(4,15)}$	$\frac{\sum (Y - \bar{Y})^2}{p} \bigg/ \frac{\sum (Y - \hat{Y})^2}{N - p - 1}$	$F_{(test)} > 2.09$	34.94	Passed	
<b>Random model</b>					
$\bar{R}_r$	an average of the correlation coefficient for randomized data	$\bar{R} < 0.5$	0.4343	passed	(Tropsha, 2010)
$\bar{R}_r^2$	an average of determination coefficient for randomized data	$\bar{R}_r^2 < 0.5$	0.2445	Passed	
$\bar{Q}_r^2$	an average of leave one out cross-validated determination coefficient for randomized data	$\bar{Q}_r^2 < 0.5$	-0.7143	Passed	
$^cR_p^2$	$R^2 \times \left( 1 - \sqrt{ R^2 - \bar{R}_r^2 } \right)$	$^cR_p^2 > 0.6$	0.7542	Passed	(Roy, 2007)
<b>External validation</b>					
$R_{Pred}^2$	$1 - \frac{\sum (Y_{ext} - \hat{Y}_{ext})^2}{\sum (Y_{ext} - \bar{Y})^2}$	$R_{pred}^2 > 0.6$	0.7482	Passed	

**Table 2 – List of some descriptors used in the QSAR optimization model**

S/NO	Descriptors symbols	Name of descriptor(s)	Class
1	AATS7s	Average Broto-Moreau autocorrelation - lag 7 / weighted by I-state	2D
2	TDB9e	3D topological distance based autocorrelation - lag 9 / weighted by Sanderson electro negativities	3D
3	RDF90i	Radial distribution function - 090 / weighted by relative first ionization potential	3D

**Table 3 – Pearson’s correlation and statistics for descriptor used in the QSAR model**

Inter- correlation			Statistics		
AATS7s	TDB9e	RDF90i	P- Value (Confidence interval)	VIF	Mean Effect (ME)
AATS7s	1		0.00043	1.232	-
TDB9e	-	1	0.00032	2.422	0.543
	0.19822			1	3
RDF90i	0.23814	-	0.00012	2.043	0.761
		0.45338		2	2

**Table 4: Y- Randomization Parameter test**

Model	R	R <sup>2</sup>	Q <sup>2</sup>
Original	0.9132	0.8751	0.8228
Random 1	0.3454	0.1193	-1.0841
Random 2	0.4868	0.2370	-1.0985
Random 3	0.4408	0.1943	-0.9815
Random 4	0.5575	0.3108	-0.5503
Random 5	0.2957	0.0874	-1.1088
Random 6	0.5562	0.3093	-0.7285
Random 7	0.7724	0.5966	0.0328
Random 8	0.2752	0.0757	-1.1166
Random 9	0.74823	0.5598	-0.0362
Random 10	0.5557	0.3088	-0.4448
<b>Random Models Parameters</b>			
Average r :	0.4343		
Average r <sup>2</sup> :	0.2445		
Average Q <sup>2</sup> :	-0.7134		

### 3.2 Interpretation of selected descriptors

**AATS7s** is Average Broto-Moreau autocorrelation - lag 7 / weighted by I-state auto-correlation descriptor. It's based on spatial dependent autocorrelation function which measures the strength of the relationship between observations (atomic or molecular properties) and space separating them (lag). This descriptor is obtained by taking the molecule atoms as the set of discrete points in space and an atomic property as the function evaluated at those points. When this descriptor is calculated on molecular graph, the lag coincides with the topological distance between any pair of the vertices. AATS7s is defined on the molecular graphs using atomic masses ( $m$ ), Sanderson electronegativity ( $e$ ) and inductive effect of pairs of atoms 7 bond apart as the weighting scheme. These observations suggested that atomic masses and electronic distribution of the atoms that made up the molecule had significant effect on the anti-tubercular activity of the dataset. In addition, the signs of the regression coefficients for each descriptor indicated the direction of influence of the descriptors in the models such that, positive regression coefficient associated to a descriptor will augment the activity profile of a compound while the negative coefficient will diminish the activity of the compound.

**TDB9e** (3D topological distance based autocorrelation - lag 9 / weighted by Sanderson electro negativities). It is positively correlated to the anticonvulsant activity meaning that increase in its value augments the activity of the studied compounds.

The descriptor measures the strength of the connection between atomic charges 9 bonds apart. The number of ring in the molecular system tends to increase the values of this descriptor as observed for molecules. This may be due to increase in the amount of  $\pi$ -electrons in the molecular system bringing about increase in the charge difference between atoms 9-bonds apart. The positive mean effect indicates a positive impact on the activity of the inhibitory compounds, which means increasing the value of this descriptor, produces higher activity of these compounds.

**RDF90i** is a radial distribution function at 2.5 and 7.0 inter-atomic distance weighted by atomic masses. The radial distribution function is probability distribution to find an atom in a spherical volume of radius. RDF descriptors are independent of the size and rotation of the entire molecule. They describe the steric hindrance or the structure-activity properties of a molecule. The RDF descriptor provides valuable information about the bond distances, ring types, planar and non-planar systems, and atom types. The presence of these descriptors in the model suggested the occurrence of a linear relationship between anti-tubercular activity and the 3D molecular distribution of atomic masses in the molecules calculated at radius of 2.0 Å and 7.0 Å from the geometrical centers of each molecule. RDF90i with positive mean effect (MF) indicates positive impact on the activity while RDF110s with negative mean effect (MF) indicates negative contribution on the activity.

Plot of predicted activity against experimental activity of training and test set where shown in Figure 1 and Figure 2 respectively. The  $R^2$  value of 0.9134 for training set and  $R^2$  value of 0.7482 for test set reported in this study was in agreement with Genetic Function Approximation (GFA) derived  $R^2$  value reported in Table 1. This confirms the robustness and reliability of the model. Plot of standardized residual versus

experimental activity shown in Figure 3 indicates that there was no systematic error in the model built as the spread of standardized residual values were on both sides of zero.

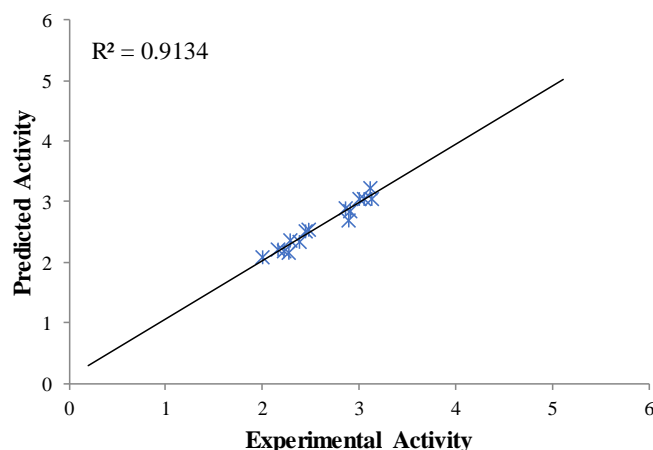


Figure 1 – Plot of predicted activity against experimental activity of training set

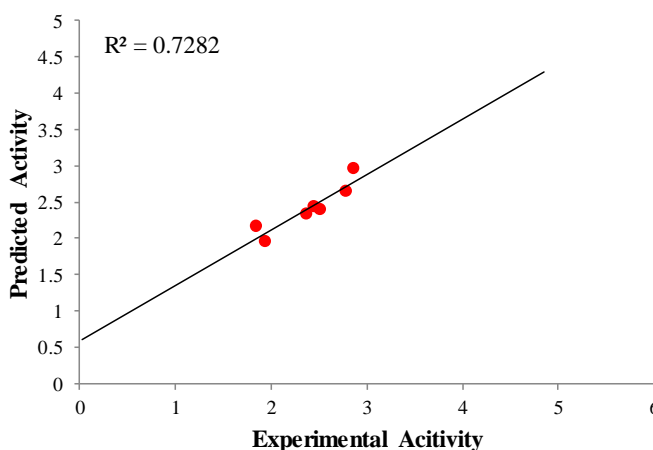


Figure 2 – Plot of predicted activity against experimental activity of test set

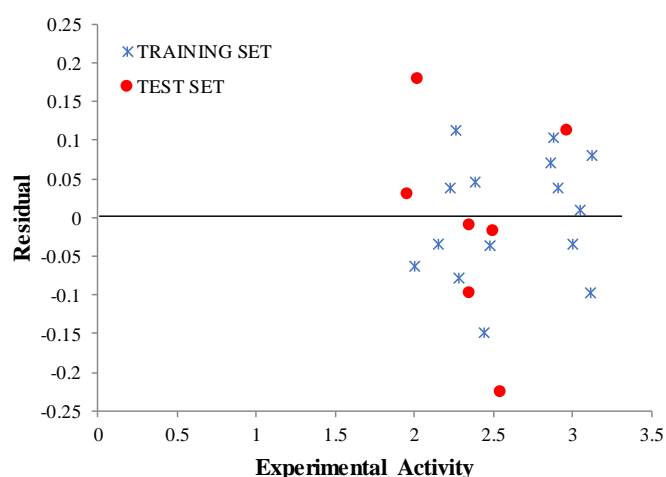


Figure 3: Plot of residual values versus experimental activity

The leverage values for the entire compounds in the dataset were plotted against their standardized residual values leading to discovery of outliers and influential compound in the models. The Williams plot of the standardized residuals versus the leverage value is shown in Figure 4. From our result it is

evident that all the compounds were within the square area  $\pm 3$  of standardized cross-validated residual produced by the model. Therefore no compound is said to be an outlier. However, only compound (number 4) is said to be an influential compound since its leverage value is greater than the warning leverage ( $h^* = 0.80$ ). This was attributed to difference in its molecular structure compared to other compounds in the dataset.

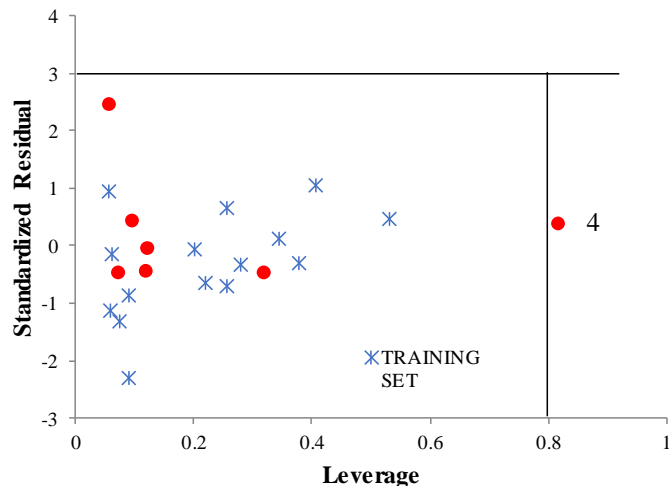


Figure 4: Plot of Standardized residual activity versus leverage

#### 4. CONCLUSION

QSAR analysis on a series of 1,2,4-Triazole derivatives was carried out using the GFA technique. The internal and external validation test for the QSAR model generated was in agreement with recommended value of validation parameters for a generally acceptable QSAR model. Thus the descriptors; AATS7s, TDB9e and RDF90i in the built model are important descriptors to determine the activity of the compounds to function as effective mycobacterium tuberculosis inhibitors.

This knowledge can be used for designing more effective chemical entities and may also provide important insights into structural variants leading to the development of novel tuberculosis inhibitors.

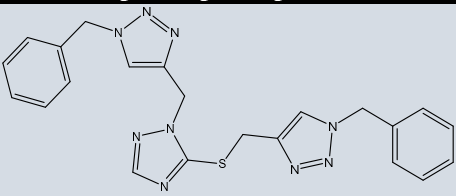
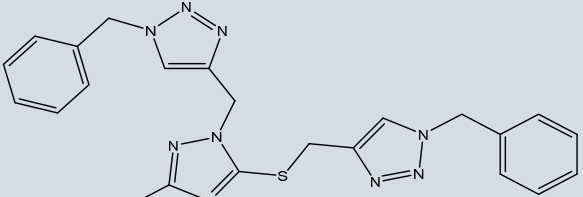
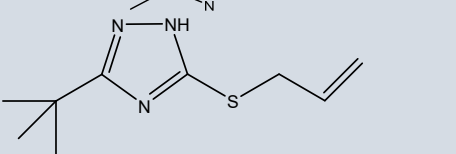
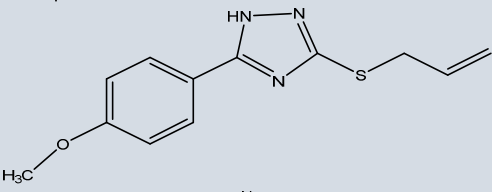
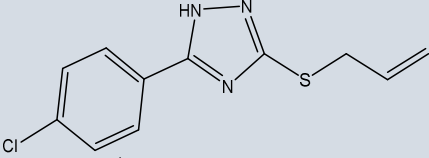
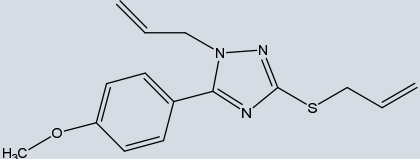
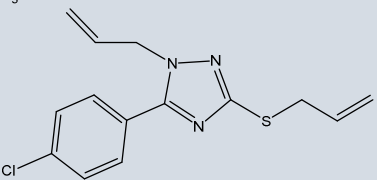
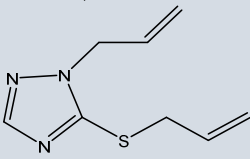
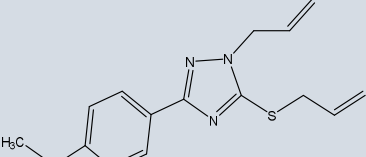
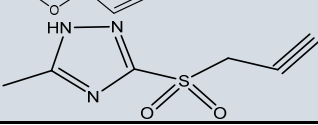
#### REFERENCES

- ABIDEEN, P. S.; CHANDRASEKARAN, K.; UMA MAHESWARAN, V. A.; KALAISELVAN, V. Implementation of self reporting pharmacovigilance in anti tubercular therapy using knowledge based approach. *Journal of Pharmacovigilance*, v.1, p.20-34, 2013.
- AFANTITIS, A.; MELAGRAKI, G.; SARIMVEIS, H.; KOUTENTIS, P. A.; MARKOPOULOS, J.; IGGLESSI-MARKOPOULOU, O. A novel QSAR model for predicting induction of apoptosis by 4-aryl-4H-chromenes. *Bioorganic and Medicinal Chemistry*, v.14, p.6686-6694, 2006.
- AZIZ, M. A.; WRIGHT, A.; LASZLO, A.; DE MUYNCK, A.; PORTAELS, F.; VAN DEUN, A.; RAVIGLIONE, M. WHO/International Union Against Tuberculosis And Lung Disease Global Project on Anti-tuberculosis Drug Resistance Surveillance. Epidemiology of antituberculosis drug resistance (the Global Project on Anti-tuberculosis Drug Resistance Surveillance): an upd. *Lancet*, v.368, p. 2142-2154, 2006.
- BALABANOVA, Y.; RUDDY, M.; HUBB, J.; YATES, M.; MALOMANOVA, N.; FEDORIN, I.; DROBNIEWSKI, F. Multidrug-resistant tuberculosis in Russia: clinical characteristics, analysis of second-line drug resistance and development of standardized therapy. *European Journal of Clinical Microbiology and Infectious Diseases*, v. 24, p. 36-139, 2005.
- CHAKRABORTI, A. K.; GOPALAKRISHNAN, B.; SOBHIA, M. E.; MALDE, A. 3D-QSAR studies of indole derivatives as phosphodiesterase IV inhibitors. *European Journal of Medicinal Chemistry*, v.38, p. 975-982, 2003.
- CRAMER, R. D., PATTERSON, D. E., & BUNCE, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *Journal of the American Chemical Society*, v. 110, p. 5959-5967, 1998.
- IBEZIM, E. C.; DUCHOWICZ, P. R.; IBEZIM, N. E.; MULLEN, L. M. A.; ONYISHI, I. V.; BROWN, S. A.; CASTRO, E. A. Computer-aided linear modeling employing QSAR for drug discovery. *Scientific Research and Essays*, v.4, p.1559-1564, 2009.
- JHAMB, S. S.; GOYAL, A.; SINGH, P. P. Determination of the activity of standard anti-tuberculosis drugs against intramacrophage Mycobacterium tuberculosis, in vitro: MGIT 960 as a viable alternative for BACTEC 460. *Brazilian Journal of Infectious Diseases*, v.18, p. 336-340, 2014.
- KENNARD, R. W.; STONE, L. A. Computer aided design of experiments. *Technometrics*, v. 11, p. 137-148, 1969.
- KHALED, K. F. Modeling corrosion inhibition of iron in acid medium by genetic function approximation method: A QSAR model. *Corrosion Science*, v. 53, p.3457-3465, 2011.
- LÖNNROTH, K.; CASTRO, K. G.; CHAKAYA, J. M.; CHAUHAN, L. S.; FLOYD, K.; GLAZIOU, P.; RAVIGLIONE, M. C. Tuberculosis control and elimination 2010-50: cure, care, and social development. *The Lancet*, v.375, p. 1814-1829, 2010.
- MELAGRAKI, G.; AFANTITIS, A.; MAKRIDIMA, K.; SARIMVEIS, H.; IGGLESSI-MARKOPOULOU, O. Prediction of toxicity using a novel RBF neural network training methodology. *Journal of Molecular Modeling*, v.12, p.297-305, 2006.
- MINOVSKI, N.; ŽUPERL, Š.; DRGAN, V.; NOVIČ, M. Assessment of applicability domain for multivariate counter-propagation artificial neural network predictive models by minimum Euclidean distance space analysis: a case study. *Analytica Chimica Acta*, v. 759, p.28-42, 2013.
- ROGERS, D.; & HOPFINGER, A. J. Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. *Journal of Chemical Information and Computer Sciences*, v.34, p.854-866, 1994.
- ROY, K. On some aspects of validation of predictive quantitative structure-activity relationship models. *Expert Opinion on Drug Discovery*, v.2, p.1567-1577, 2007.

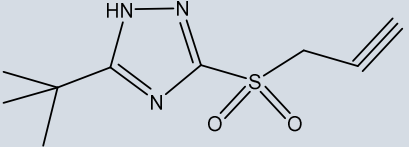
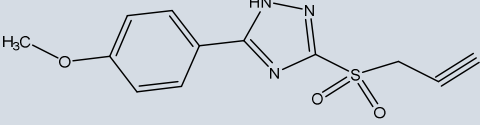
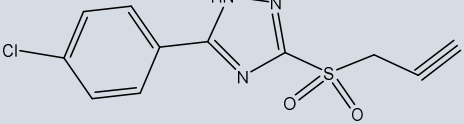
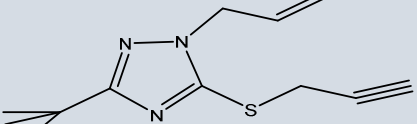
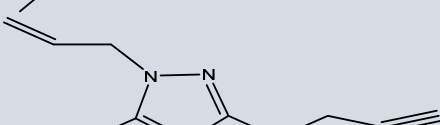
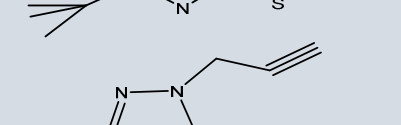
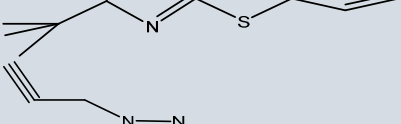
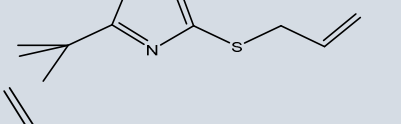
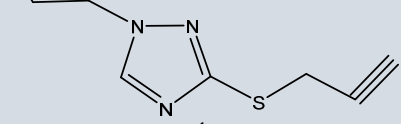
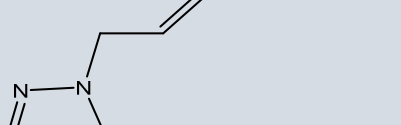
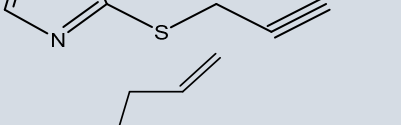
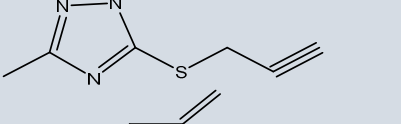
- SARKAR, D.; DESHPANDE, S. R.; MAYBHATE, S. P.; LIKHITE, A. P.; SARKAR, S.; KHAN, A.; CHAVAN, S. R. 1, 2, 4- derivatives and their anti-microbial activity. Google Patents.2016.
- TROPSHA, A. Best practices for QSAR model development, validation, and exploitation. *Molecular Informatics*, v. 29, p. 476–488, 2010.
- TROPSHA, A.; GRAMATICA, P.; GOMBAR, V. K. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *Molecular Informatics*, v.22, p.69–77, 2003.
- WU, W.; WALCZAK, B.; MASSART, D. L.; HEUERDING, S.; ERNI, F.; LAST, I. R.; PREBBLE, K. A. Artificial neural networks in classification of NIR spectral data: design of the training set. *Chemometrics and Intelligent Laboratory Systems*, v. 33, p.35–46, 1996.
- YAKAR, A.; YAKAR, F.; YILDIZ, Z. Isoniazid-and rifampicin- induced thrombocytopenia. *Multidisciplinary Respiratory Medicine*, v.8, 13, 2013.

## APENDIX A

Table A1 – Molecular structure of 1, 2, 4-Triazole derivatives and their activities as potent anti-mycobacterium tuberculosis.

S/N	Molecules superscript a represent the test set	Experimental Activity (pBA)	Predicted Activity (pBA)	Residual
1 <sup>a</sup>		6.3456	6.339078	0.006522
2		7.4134	7.50969	-0.09629
3		6.4171	6.42649	-0.00939
4		7.6397	7.459252	0.180448
5		8.0899	8.059256	0.030644
6 <sup>a</sup>		7.366	7.42033	-0.05433
7		7.0123	6.89972	0.11258
8 <sup>a</sup>		6.5267	6.75046	-0.22376
9 <sup>a</sup>		7.3233	7.326721	-0.00342
10		7.3279	7.42522	-0.09732



11		6.8568	6.776661	0.080139
12		7.3079	7.270438	0.037462
13		7.314	7.248722	0.065278
14		8.5854	8.66399	-0.07859
15		8.0615	8.12506	-0.06356
16		8.0615	7.972504	0.088996
17		6.8494	6.99939	-0.14999
18 <sup>a</sup>		7.9432	7.934651	0.008549
19		7.4535	7.444966	0.008534
20		7.9759	8.01127	-0.03537
21 <sup>a</sup>		7.9759	8.05262	-0.07672
22 <sup>a</sup>		7.9294	7.852469	0.076931

Where superscript **a** represent the test set