**Turkish Computational and Theoretical Chemistry**

*Turkish Comp Theo Chem (TC&TC)*

***Volume(Issue): 2(2) – Year: 2018 – Pages: 16-27***

***ISSN: 2587-1722 / e-ISSN: 2602-3237***

***Received:*** 21.02.2018 ***Accepted:*** 28.08.2018 *Research Article*

# *QSAR Studies of amino-pyrimidine derivatives as Mycobacterium tuberculosis Protein Kinase B inhibitors*

**Saida Khamouli[a], Salah Belaidi[1,a], Houmam Belaidi[a,b], Lotfi Belkhiri[1,c]**

[a] Group of Computational and Medicinal Chemistry, Laboratory of Molecular Chemistry and environment, Department of Chemistry , University of Biskra, BP 145 Biskra 07000, Algeria.

[b] Institut des Sciences Chimiques de Rennes, UMR 6226 CNRS-Université de Rennes 1, Campus de Beaulieu, 35042, Rennes Cedex, France

[c] URCHEMS, Department of Chemistry, University of Constantine 1, 25017, Constantine, Algeria

**Abstract:** Quantitative structure activity relationship (QSAR) analysis was applied to a series of amino-pyrimidine derivatives as PknB inhibitors using a combination of various physicochemical and quantum descriptors. A multiple linear regression (MLR) procedure was used to model the relationships between molecular descriptors and the chemotherapeutic activity of the amino-pyrimidine derivatives. Good agreement between experimental and predicted activity values, obtained in the validation procedure, indicated the good quality of the derived QSAR model. The best QSAR model developed show a good predictive correlation coefficient R2= 0.973 and external predictive ability of prediction R2=0.778 was developed by MLR. The proposed model has good robustness and predictability when verified by internal and external validation.

***Keywords:*** *QSAR, amino-pyrimidine, PknB, MLR.*

## 1. Introduction

Tuberculosis (TB) is the primary cause of mortality due to an infectious disease in the world today. The causative agent of tuberculosis is the intracellular pathogen M. tuberculosis [1]. Recent studies have focused on finding new pathways vulnerable to inhibition by small molecules and previously unexploited by drug discovery efforts [2]. The inhibition of signaling pathways both in M. tuberculosis and the host may yield new classes of drug targets and a large amount of recent studies are focused on developing this further [3].

Target based drug discovery, in which there is in vitro high throughput screening of a large number of small molecules against a validated target, has been used on a number of occasions to search for new anti-tuberculosis agents. We sought to find inhibitors of an essential mycobacterium tuberculosis serine/threonine protein kinase, (PknB) [4].

Novel molecules such as chalcones, thiophene, pyrimidine derivatives was synthesis exhibit antimycobacterial activity against other species of Mycobacterium [5-8].

Despite the fact that death from TB is often preventable, the rapid increase of multidrug-resistant tuberculosis and extensively drug-resistant tuberculosis has resulted in an urgent need to develop new drug targets for Mycobacterium tuberculosis (Mtb) [9, 10].

The discovery and development of new anti-TB therapeutics is widely recognized as one of the major global health emergencies, yet it is also a major pharmaceutical challenge. A theoretical technique such as quantum chemical descriptors have been extensively used in Quantitative Structure-Activity Relationship (QSAR) [11-14] to predict the physiological and biological properties of compounds understudy. Whether these compounds (amino-pyrimidine derivatives)

[1] Corresponding authors

*E-mail*: prof.belaidi@gmail.com (Salah Belaidi) and lotfi.belkhiri@umc.edu.dz (Lotfi Belkhiri)

provide structural precedence and may lead to the generation of novel anti-TB therapeutics.

The main goal of QSAR studies is the development of predictive models which can be used in computer-aided drug discovery and design for prediction of activities or properties of new compounds (i.e., those included in chemical databases or combinatorial libraries). Prior to using the model for external predictions, its predictive power should be established and validated. Thus, model validation has become a standard (and in some laboratories such as ours, mandatory) part of QSAR modeling. In the past, model validation was performed using leave one out (LOO) cross-validation. LOO cross-validation has been determined to yield an overoptimistic estimate of predictive ability (if model feature selection is not redone for each hold out) [15].

External validation is now a "must have" tool for evaluating the reliability of QSAR models [16]. In this procedure, typically the overall set is randomly divided into a training set and a test set. QSAR models are developed based on the training set and then are used to make predictions for the test set. The advantage of this approach is that the test set compounds are "unknown" to the models since they are excluded from the model development procedure, especially the variable selection.

## 2. Materials and methods

### 2.1. Data Set

A dataset of 29 amino-pyrimidine derivatives as Mycobacterium tuberculosis PknB inhibitors was selected from the literature based on the biological assay method [3, 17-18]. The structures and their inhibitory activities in pIC50 are listed in Table 1. pIC50 = - Log (IC50) .IC50 represents the compound concentration required for 50% inhibition.

The data set was divided into the two subsets, training set of 22 compounds (75.76%) and test set of 7 compounds (24.24% >20% for external validation). The test set compounds were selected manually considering the distribution of biological data and structural diversity. Training set was used to build a regression model, and the test set was used to evaluate the predictive ability of the model obtained.

Firstly, all the structures of molecules were pre-optimized using the Molecular Mechanics Force Field (MM+) method included in HyperChem release 8.08 [19]. The resulted minimized structures were further refined using the semi empirical PM3 Hamiltonian, which is implemented also in HyperChem. We chose a gradient norm limit of 0.01 kcal/Å for the geometry optimization.

In the second part, the series of the amino-pyrimidine derivatives was re-optimised by the density functional theory DFT/B3LYP at the 6-311G(d,p) basis set of level, by using Gaussian 09 software [20].Further, the regression analysis was performed using the SPSS version 21 for Windows [21].

The QSAR properties module from HyperChem (8.08) was used to calculate: molar polarizability (Pol), the molar refractivity (MR), partition coefficient octanol/water(log P), molar volume (MV), Surface area grid (SAG) and molar weight (MW).

The Quantum Chemical descriptors: dipole moment (DM), HOMO-LUMO energy gap (ΔE), total energy (ET) and atomic net charges (Natural bond orbital charges NBO )(qN1, qC2, qN3, qC5 and qN6)were computed using Gaussian 09 software.

### 2.2. Multiple linear regression (MLR) analysis

Multiple linear regression is the standard method for multivariate data analysis. It is also called as Ordinary Least Squares regression (OLS). This method of regression estimates the values of the regression coefficients by applying least squares curve fitting method. For getting reliable results, dataset having typically 5 times as many data points (molecules) as independent variables (descriptors) is required. The regression equation takes the form:

$$Y = a0+a1*X1 + a2*X2 + a3*X3 +.....+an*Xn \quad (1)$$

Where Y is the response or dependent variable, X1, X2, …, Xn are descriptors (features or independent variables) present in the model with the corresponding regression coefficients a1, a2, …, an, respectively, and a0 is the constant term of the model.

### 2.3. Validation Methods for QSAR Models

For the validation of the predictive power of a QSAR model, two basic principles (internal validation and external validation)

#### 2.3.1. Internal validation

Internal model validation was carried out using leave-one-out (LOO-$R^2_{CV}(Q^2)$) method. For

calculating $R^2_{cv}(Q^2)$, each sample in the training set was eliminated once and the activity of the eliminated sample was predicted by using the model developed by the remaining samples. LOO-$R^2_{cv}(Q^2)$ calculated [22]according to the below formula:

$$R^2_{CV}(\mathrm{Q2}) = 1 - \frac{\sum(Ypred-Yobs)^2}{\sum(Ypred-Ymean)^2} > 0.5 \quad (2)$$

In Eq. (2), $Y_{pread}$ and $Y_{obs}$ indicate predicted and observed activity values accordingly and $Y_{mean}$ signify mean activity value. A model is considered acceptable when the value of $R^2_{CV}(Q^2)$ exceeds 0.5.

### 2.3.2. External validation

According to Tropsha et al. [23], for considering the validity of the developed models, in addition to the internal validation, the models should be externally validated using the test set compounds. According to their study, the following criteria could be considered as of acceptable predictability:

$$R^2_{pred} = 1 - \frac{\sum(Ypred-Yobs)^2}{\sum(Ypred-Ymean(train))^2} > 0.6 \quad (3)$$

where $R^2_{pred}$ is squared correlation coefficient between the experimental and predicted biological activities of dataset compounds (training and test sets).

Ypred and Yobs indicate predicted and observed activity values for the test set and Ymean(train) indicates mean activity value of the training set. For the predictive QSAR model.

$$\frac{(R^2-R^2_0)}{R^2}<0.1 \quad \text{or} \quad \frac{(R^2-R'^2_0)}{R^2}<0.1 \quad (4)$$

$$|R^2_0 - R'^2_0| < 0.3 \quad (5)$$

where $R^2$ is determination coefficient between the experimental and predicted pIC50 values of test set compounds with intercept.$R^2_0$ and $R'^2_0$ are determination coefficients of predicted versus experimental and experimental versus predicted biological activity values in which their regression line must pass through the origin, respectively.

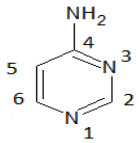
$$0.85 \leq K \leq 1.15 \quad \text{or} \quad 0.85 \leq K' \leq 1.15 \quad (6)$$

The slope of regression line through the origin which mentioned in the above was indicated by K and K', respectively.

## 3. Results and Discussion

In an attempt to determine the role of structural features of compounds, which appears to have an effect the antitubercular activity, QSAR models were generated. The structures of studied compounds are shown in Table 1. , and descriptors are used in Table 2. To establish the statistical correlation, the physicochemical descriptors were taken as independent variables and antitubercular activity as dependent variable

**Table 1.** Chemical structure, experimental and predicted activities of the molecules under study.



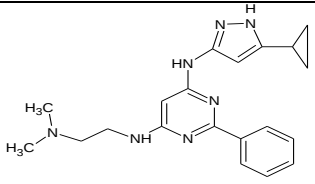
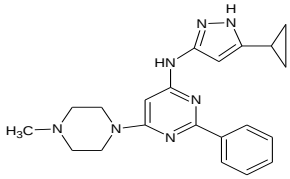

| No | Compound | pIC50 Exp. | pIC50 Pred. | Res. |
|---|---|---|---|---|
| 1 |  | 7.06 | 7.118 | -0.058 |
| 2 |  | 6.92 | 6.873 | 0.046 |

**Table 1.** Continued

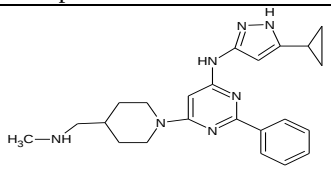
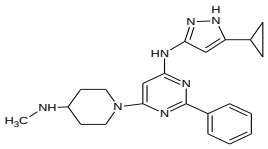
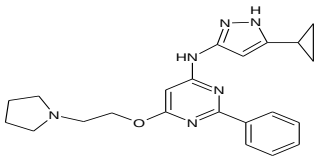
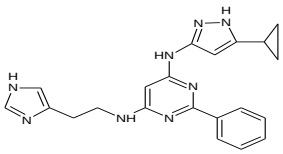
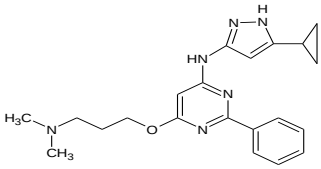
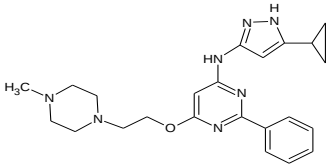
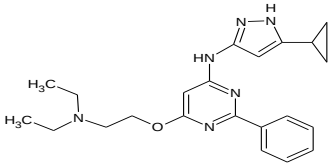
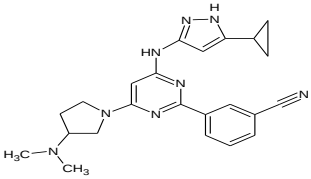
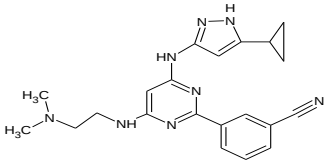
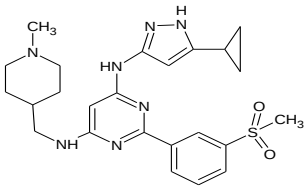
| No | Compound | pIC50 Exp. | pIC50 Pred. | Res. |
|---|---|---|---|---|
| $3^T$ |  | 7.14 | 6.865 | 0.274 |
| $4^T$ |  | 7.28 | 7.240 | 0.040 |
| 5 |  | 7.25 | 7.165 | 0.084 |
| 6 |  | 6.46 | 6.491 | -0.031 |
| 7 |  | 7.26 | 7.204 | 0.055 |
| 8 |  | 7.02 | 7.216 | -0.196 |
| 9 |  | 7.04 | 6.951 | 0.088 |
| 10 |  | 7.25 | 7.182 | 0.067 |
| 11 |  | 7.41 | 7.372 | 0.037 |
| 12 |  | 7.40 | 7.402 | -0.002 |

**Table 1.** Continued

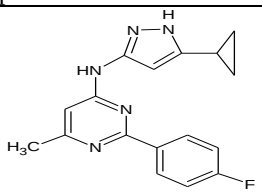
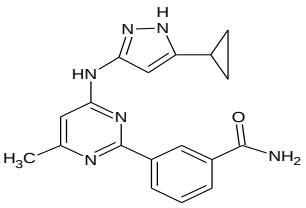
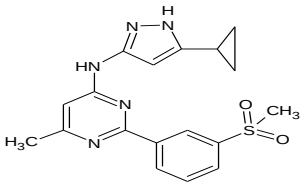
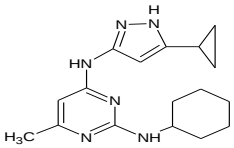
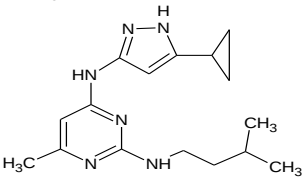
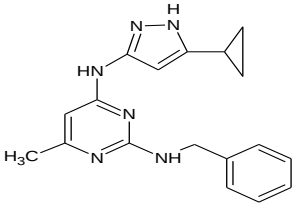
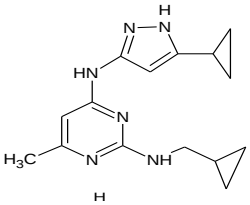
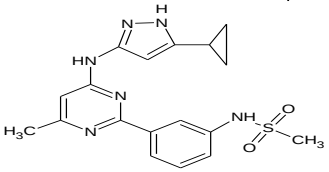
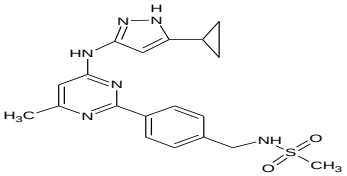
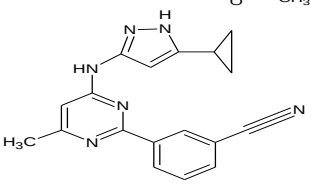
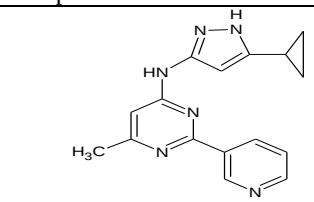
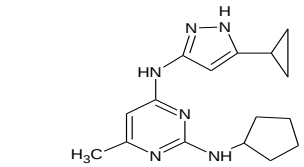
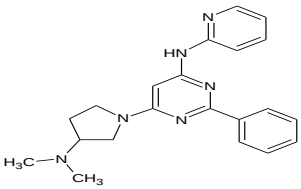
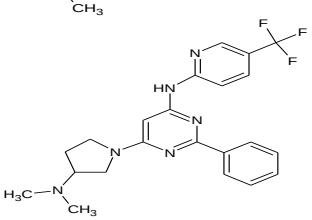
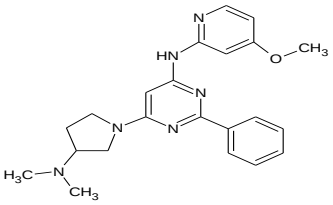
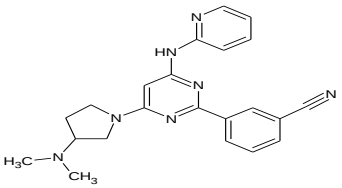
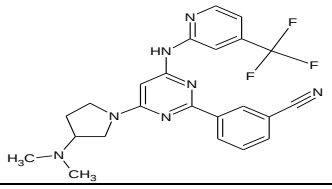
| No | Compound | pIC50 Exp. | pIC50 Pred. | Res. |
|---|---|---|---|---|
| 13 |  | 6.43 | 6.618 | -0.188 |
| 14 |  | 6.42 | 6.354 | 0.064 |
| 15 |  | 7.07 | 6.951 | 0.118 |
| 16[T] |  | 7.08 | 6.327 | 0.752 |
| 17 |  | 6.45 | 6.389 | 0.061 |
| 18[T] |  | 6.15 | 6.304 | -0.154 |
| 19 |  | 6.66 | 6.712 | -0.053 |
| 20 |  | 6.41 | 6.432 | -0.022 |
| 21 |  | 6.62 | 6.644 | -0.025 |
| 22 |  | 7.06 | 7.047 | 0.013 |

**Table 1.** Continued

| No | Compound | pIC50 Exp. | pIC50 Pred. | Res. |
|---|---|---|---|---|
| 23[T] |  | 6.69 | 6.649 | 0.041 |
| 24[T] |  | 6.94 | 7.225 | -0.285 |
| 25[T] |  | 5.21 | 5.147 | 0.062 |
| 26 |  | 6.28 | 6.313 | -0.033 |
| 27 |  | 5.95 | 5.941 | 0.009 |
| 28 |  | 5.70 | 5.683 | 0.016 |
| 29 |  | 7.40 | 7.451 | -0.051 |

Exp: Experimental activity, Pred: Predicted activity, T: Test Set, Res: Residual

**Table 2.**Values of molecular descriptors used in the regression analysis.

| Comp number | MW (amu) | SAG ($Å^2$) | MV ($Å^3$) | Pol ($Å^3$) | MR ($Å^3$) | HE (Kcal/mol) | LogP | MD Debye |
|---|---|---|---|---|---|---|---|---|
| 1 | 363.47 | 676.89 | 1137.04 | 41.62 | 113.02 | -8.87 | 1.81 | 4.74 |
| 2 | 375.48 | 666.97 | 1130.58 | 42.69 | 116.35 | -7.28 | 1.99 | 5.71 |
| 3 | 403.53 | 727.93 | 1234.55 | 46.36 | 125.27 | -7.95 | 2.23 | 2.58 |
| 4 | 389.50 | 685.27 | 1175.67 | 4452 | 120.46 | -7.93 | 1.81 | 1.74 |
| 5 | 390.49 | 692.41 | 1182.78 | 43.81 | 118.62 | -7.51 | 2.45 | 2.50 |
| 6 | 386.46 | 648.19 | 1128.08 | 43.43 | 118.95 | -17.6 | 0.42 | 2.00 |
| 7 | 378.48 | 700.00 | 1178.93 | 42.75 | 115.94 | -7.95 | 2.18 | 3.72 |
| 8 | 419.53 | 720.61 | 1250.68 | 46.99 | 127.13 | -7.72 | 1.93 | 2.07 |
| 9 | 392.50 | 721.55 | 1221.92 | 44.58 | 120.57 | -7.45 | 2.81 | 4.16 |
| 10 | 414.51 | 715.62 | 1240.43 | 46.37 | 125.86 | -9.83 | 1.84 | 6.11 |
| 11 | 388.47 | 690.78 | 1175.96 | 43.48 | 118.00 | -9.81 | 1.53 | 5.90 |
| 12 | 481.62 | 791.24 | 1369.02 | 49.33 | 138.94 | -12.26 | 0.64 | 9.65 |
| 13 | 309.35 | 561.65 | 917.18 | 33.33 | 91.97 | -7.37 | 2.06 | 3.94 |
| 14 | 334.38 | 580.68 | 982.60 | 36.69 | 99.66 | -11.52 | 1.18 | 1.45 |
| 15 | 369.44 | 604.52 | 1035.66 | 36.39 | 106.69 | -9.22 | -0.12 | 4.28 |
| 16 | 312.42 | 604.63 | 996.87 | 35.34 | 92.11 | -6.44 | 3.41 | 2.68 |
| 17 | 300.41 | 593.65 | 985.85 | 34.28 | 89.52 | -6.57 | 3.36 | 2.83 |
| 18 | 320.40 | 603.62 | 999.47 | 36.60 | 99.85 | -9.38 | 2.84 | 2.84 |
| 19 | 330.39 | 617.61 | 1115.49 | 37.10 | 83.04 | -7.44 | 2.72 | 2.94 |
| 20 | 384.46 | 645.49 | 1080.42 | 37.74 | 109.99 | -12.29 | -0.16 | 2.30 |
| 21 | 398.48 | 675.55 | 1141.48 | 39.58 | 112.97 | -11.81 | 1.54 | 9.70 |
| 22 | 316.37 | 588.60 | 962.93 | 35.27 | 96.82 | -13.21 | 2.39 | 8.99 |
| 23 | 292.34 | 552.48 | 895.67 | 32.71 | 88.34 | -9.69 | 1.32 | 8.39 |
| 24 | 293.39 | 571.88 | 943.99 | 33.51 | 88.70 | -6.85 | 1.83 | 2.95 |
| 25 | 360.46 | 637.86 | 1092.89 | 41.92 | 114.92 | -9.00 | 2.30 | 3.60 |
| 26 | 428.46 | 682.82 | 1173.21 | 43.48 | 120.33 | -4.58 | 3.25 | 7.75 |
| 27 | 390.49 | 688.88 | 1173.45 | 44.39 | 121.30 | -7.30 | 1.31 | 5.34 |
| 28 | 385.47 | 669.46 | 1153.85 | 43.77 | 116.52 | -10.49 | 3.15 | 5.61 |
| 29 | 453.47 | 694.49 | 1209.04 | 45.33 | 125.12 | -9.85 | 2.59 | 8.48 |

The MLR QSAR model was developed by applying the selected molecular descriptors to the "active compounds" of the training data (22 compounds).

The best model was selected on the basis of statistical parameters via observed squared correlation coefficient ($R^2$>0.6) which is a relative measure of quality of fit. Standard error of estimate (SEE < 0.3) representing absolute measure of quality of fit, Fischer's value (F), is the Fisher ratio, reflects the ratio of the variance explained by the model and the variance due to the error in the regression. High values of the *F* -test indicate that the model is statistically significant [24].

Pearson's correlation matrix has been performed on all descriptors by using SPSS statistics 21 Software. The analysis of the matrix revealed ten descriptors for the development of MLR model.

The correlation between the biological activities and descriptors expressed by the following relation:

$$pIC50 = -50.585 - 0.002SAG + 0.065H - 0.260\,Log\,P + 0.128MD + .001ET - 49.366qN1 - 11.471qC2 + 14.051qN3 + 107.556qC4 + 11.553qC5 \quad (7)$$

*N= 22 R=0.986 SEE= 0.110 F = 39.341 Q = 8.963*

where N is the number of compounds (training set).The model shows a good correlation coefficient (*R*) of 0.986 between descriptors (Log *P*, SAG, HE, MD, ET, $qN_1$, $qC_2$, qN3, $qC_5$ and $qC_4$ ).

In this model the values of $qC_4$, $qN_3$, $qC_5$, ET,MD and HE suggest that the activity increases with the increase of these descriptors values, on the other hand the activity decreases with increasing the values of $qN_1$, $qC_2$, log P and SAG. Since they have a negative value in this equation (7).

**Table 2.** Continued.

| Compound | ET a.u | ΔE a.u | qN1 | qC2 | qN3 | qC4 | qC5 | qC6 |
|---|---|---|---|---|---|---|---|---|
| 1 | -1160.695 | 0.234 | -0.581 | 0.469 | -0.558 | 0.460 | -0.406 | 0.455 |
| 2 | -1198.813 | 0.162 | -0.593 | 0.466 | -0.585 | 0.453 | -0.399 | 0.469 |
| 3 | -1277.461 | 0.234 | -0.588 | 0.474 | -0.592 | 0.460 | -0.396 | 0.463 |
| 4 | -1238.144 | 0.241 | -0.593 | 0.475 | -0.592 | 0.449 | -0.394 | 0.463 |
| 5 | -1258.010 | 0.239 | -0.604 | 0.476 | -0.577 | 0.454 | -0.375 | 0.591 |
| 6 | -1251.789 | 0.242 | -0.597 | 0.474 | -0.592 | 0.455 | -0.388 | 0.455 |
| 7 | -1219.888 | 0.237 | -0.597 | 0.472 | -0.542 | 0.451 | -0.378 | 0.593 |
| 8 | -1352.677 | 0.240 | -0.608 | 0.475 | -0.578 | 0.453 | -0.374 | 0.591 |
| 9 | -1259.211 | 0.237 | -0.595 | 0.473 | -0.542 | 0.451 | -0.378 | 0.594 |
| 10 | -1330.400 | 0.242 | -0.592 | 0.467 | -0.593 | 0.459 | -0.391 | 0.460 |
| 11 | -1252.970 | 0.247 | -0.583 | 0.465 | -0.592 | 0.464 | -0.399 | 0.456 |
| 12 | -1865.431 | 0.251 | -0.601 | 0.466 | -0.579 | 0.455 | -0.398 | 0.464 |
| 13 | -1031.279 | 0.244 | -0.557 | 0.468 | -0.570 | 0.458 | -0.336 | 0.282 |
| 14 | -1100.770 | 0.052 | -0.557 | 0.464 | -0.564 | 0.458 | -0.332 | 0.279 |
| 15 | -1519.949 | 0.244 | -0.556 | 0.461 | -0.566 | 0.458 | -0.324 | 0.276 |
| 16 | -991.024 | 0.251 | -0.613 | 0.648 | -0.611 | 0.464 | -0.382 | 0.288 |
| 17 | -952.905 | 0.252 | -0.612 | 0.645 | -0.613 | 0.464 | -0.380 | 0.290 |
| 18 | -1026.712 | 0.255 | -0.612 | 0.645 | -0.611 | 0.464 | -0.380 | 0.290 |
| 19 | -912.333 | 0.250 | -0.617 | 0.645 | -0.612 | 0.464 | -0.383 | 0.289 |
| 20 | -1575.339 | 0.253 | -0.547 | 0.460 | -0.536 | 0.460 | -0.338 | 0.279 |
| 21 | -1614.662 | 0.244 | -0.550 | 0.462 | -0.532 | 0.456 | -0.337 | 0.280 |
| 22 | -1024.280 | 0.052 | -0.556 | 0.459 | -0.557 | 0.460 | -0.349 | 0.292 |
| 23 | -948.055 | 0.245 | -0.548 | 0.453 | -0.562 | 0.454 | -0.341 | 0.291 |
| 24 | -951.687 | 0.247 | -0.620 | 0.644 | -0.612 | 0.464 | -0.382 | 0.288 |
| 25 | -1143.499 | 0.229 | -0.428 | 0.362 | -0.424 | 0.462 | -0.237 | 0.464 |
| 26 | -1480.623 | 0.226 | -0.576 | 0.471 | -0.549 | 0.451 | -0.387 | 0.463 |
| 27 | -1258.055 | 0.230 | -0.585 | 0.466 | -0.588 | 0.447 | -0.390 | 0.465 |
| 28 | -1235.762 | 0.241 | -0.579 | 0.470 | -0.598 | 0.454 | -0.384 | 0.463 |
| 29 | -1572.897 | 0.233 | -0.580 | 0.467 | -0.578 | 0.448 | -0.400 | 0.472 |

The cross-validated $R^2_{CV}$($Q^2$) was found to be very close to the value of $R^2_{CV}$ for the training set and hence this model can be termed as statistically significant. The QSAR model expressed by (Eq.7) were cross validated by the high value of $R^2_{CV}$= 0.973 obtained by leave one out (LOO).The value of $R^2_{CV}$($Q^2$) is indeed greater than 0.5which is the essential condition to qualify a QSAR model as valid [22], and RMSE of cross-validation was 0.078, which indicates reliability of the proposed model.

Low value of standard error of estimate (<0.3) indicates the accuracy of the statistical fit. All the values of the *t*-statistic are significant which confirms the significance of each descriptor.

The calculated *F* value for the generated QSAR model exceeds the tabulated *F* value by large margin as desired for a meaningful regression. Furthermore, the calculated *F* Value also determines a confidence limit superior to 95% for this model. The positive value of quality factor (Q) for QSAR model suggests its high predictive power and lack of over fitting, low standard deviation of the model demonstrates accuracy of the model.

**Table 3.** Cross-validation parameters.

| Model | PRESS | SSY | PRESS/SSY | $S_{PRESS}$ | $R^2_{cv}$ | $R^2_{adj}$ | 6PE |
|---|---|---|---|---|---|---|---|
| | 0.135 | 4.963 | 0.027 | 0.077 | 0.973 | 0.948 | 0.023 |

The present QSAR study can be successfully applied to predict inhibitor activity of amino-pyrimidine in these molecule generations. To investigate the presence of a systematic error in developing the QSAR models, in the figure 1 are shown the residuals of predicted values of the biological activity pIC50 which are plotted against the experimental values. The propagation of the residuals on both sides of zero indicates that no systematic error exists, as suggested by Jalali-Heravi and Kyani [26].

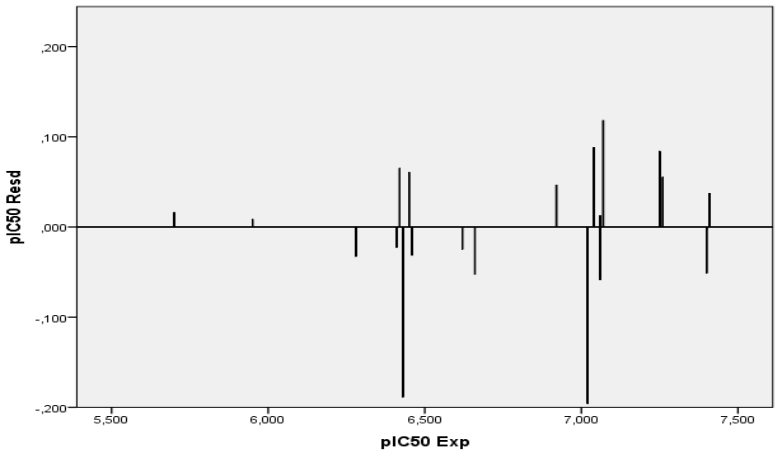



**Fig. 1.** Plots of the residual values against the experimentally observed.

In order to test the validity of the predictive power of selected MLR model (Eq.7), the leave-one out technique (LOO technique) was used. The developed models were validated by calculation of the following statistical parameters: predicted residual sum of squares (PRESS), total sum of squares deviation (SSY), the predictive error of the coefficient of correlation (PE) and cross-validated correlation coefficients ($R^2_{adj}$ and $R^2_{cv}$) (Table 3).

PRESS is an important cross-validation parameter as it is a good approximation of the real predictive error of the models. Its value being less than SSY points out that the model has a good predictive power and can be considered statically significant. The smaller PRESS value means the better model predictability. From the results reported in Table 3, this model is statistically significant. Furthermore, for reasonable QSAR model, the PRESS/SSY ratio should be lower than 0.4. [25]

The data presented in Table 3 indicate that for the developed model this ratio is 0.027. The resulted value of $R^2_{CV}$ for these QSAR model is 0.973.The high value of $R^2_{CV}$ and $R^2_{adj}$ are essential criteria for the best qualification of the QSAR model.

The predictive error of the coefficient of correlation (PE) is yet another parameter used to evaluate the predictive power of the proposed models. We have calculated the PE value of the proposed model and they are reported in Table 3. For the developed model the condition R = 0.986> 6PE is confirmed and hence, they could be considered as good predictive power. Also, for the validation of the QSAR model we went through the test set via training set. The test set contained seven compounds.

The model (Eq. 7) also passed successfully Tropsha's[23] recommended tests for predictive ability:

$R^2_{pred} = 0.778 > 0.6$

$\frac{(R^2-R_0^2)}{R^2} = -0.225 < 0.1$

$\frac{(R^2-R'^2_0)}{R^2} = -0.218 < 0.1$

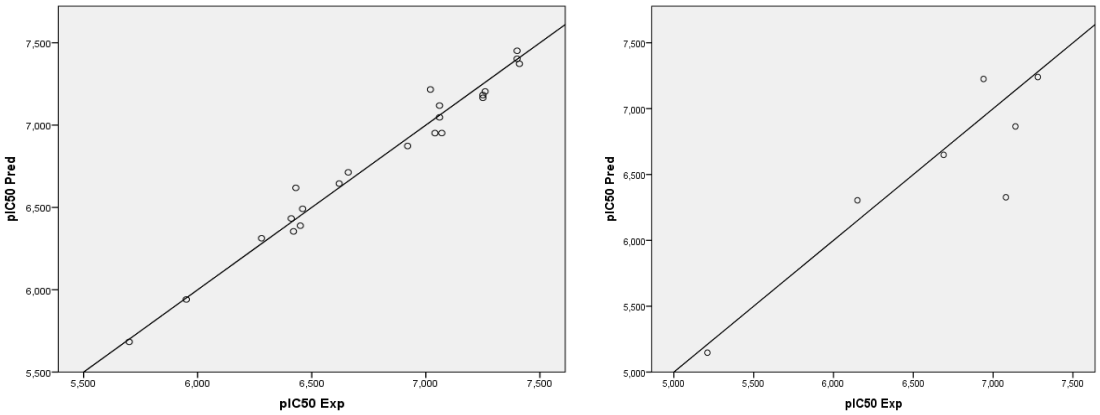
k= 1.014 k'=0.983 $(0.85 \leq k \leq 1.15)$

$|R_0^2 - R'^2_0| = 0.006 < 0.3$

These results might be considered as an indicator of good external predictability.

The built model produced good results for the training set and the test set. It is noteworthy, that the MLR equation has acceptable quality and can predict the activity of training and test set with $R^2_{CV}$ = 0.973 and $R^2_{pred}$= 0.778, respectively.

The plots of the predicted pIC50 versus the experimental pIC50, obtained by the MLR modeling, are demonstrated in Figure 2. Therefore; we conclude that the antitubercular activity is related to the physicochemical molecular descriptors and Quantum descriptors.




a b

**Fig. 2.** Data fitness plot of Best Multiple linear regression (a) for the training set, (b) for the test set

## 4. Conclusion

The QSAR analysis was conducted with a series of 29 amino-pyrimidine derivatives mycobacterium tuberculosis (PknB) inhibitors, the model depending on the (eq.7). MLR is the best produced model with very good statistical fit as evident, R = 0.986 and F=39.341.The physicochemical molecular and Quantum descriptors were found to have a key role in governing the change in activity. The model was validated using LOO cross-validation, and external test set. The robustness of the constructed model used in this study has great predictive power, as assessed by the internal and external validations. It indicates the model can be successfully applied to predict the antitubercular activity of these classes of molecules. The applicability domain served as a valuable tool to filter out ‘‘dissimilar’’ compounds. Due to its predictive ability, the proposed model could be a useful aid to the costly and time consuming experiments for determining antitubercular activit y.The proposed QSAR workflow aims to help researchers to design a novel chemistry driven molecules with desired biological activity.

## References

[1] K. McDonough, Y. Kress, B. R. Bloom, Pathogenesis of tuberculosis: interaction of Mycobacterium tuberculosis with macrophages, Infection and Immunity 61 (1993) 2763-2773.

[2] K. Rohini , P. S. Srikumar, Insights from the docking and molecular dynamics simulation of the Phosphopantetheinyl transferase (PpTt) structural model from Mycobacterium tuberculosis, Bioinformation 9(2013) 685–689.

[3] K. E. Lougheed, S. A. Osborne, B. Saxty, D. Whalley , T. Chapman, N. Bouloc , et al, Effective inhibitors of the essential kinase PknB and their potential as anti-mycobacterial agents, Tuberculosis (Edinburgh, Scotland) 91 (2011) 277-286.

[4] S. Lapenna, A. Giordano, Cell cycle kinases as therapeutic targets for cancer, Nature Reviews. Drug Discovery 8 (2009)547-566.

[5] V. U. Ahmad, A. Khan, U. F. arooq, F. Kousar, S.S. Khan, S. A. Nawaz, M. A. Abbasi , M. I. Choudhary, Three new cholinesterase-inhibiting cis-clerodanediterpenoids from Otostegialimbata , Chemical & Pharmaceutical Bulletin 53(2005)378-381.

[6] Z. N. Siddiqui , F. Farooq, Heterocyclic Silica supported sodium hydrogen sulfate ($NaHSO_4$–$SiO_2$): A novel, green catalyst for synthesis of pyrazole and pyranyl pyridine derivatives under solvent-free condition via heterocyclic β-enaminones, Journal of Molecular Catalysis A: Chemical 363-364 (2012) 451-459.

[7] Y. LiM, H.W. Xu, W.Fan, Q. Ye, X. Wang, B. Jiang, S. L. Wan, S. J. Tu ,New formal (3+3) cycloaddition of enaminones for forming tetracyclic indolo[2,3-b]quinolines under microwave irradiation ,Tetrahedron70(2014)1004-1010.

[8] S. N. Bukhari, S. G. Franzblau, I. Jantan , M. Jasamai, Current prospects of synthetic curcumin analogs and chalcone derivatives against mycobacterium tuberculosis, Medicinal Chemistry 9(2013)897-903.

[9] J. B. Nachega, R. E .Chaisson ,Tuberculosis drug resistance: a global threat, Clinical Infectious Diseases 36(2003) 24-30.

[10] M. Raviglione, I. Smith, XDR tuberculosis-implications for global public health, New England Journal Of Medicine 356 (200) 656–659.

[11] S. Belaidi, T. Salah, N. Melkemi, L. Sinha, O. Prasad , Structure Activity Relationship and Quantitative Structure-Activity Relationships Modeling of Antitrypanosomal Activities of AlkyldiamineCryptolepine Derivatives, Journal of Computational and Theoretical Nanoscience 12(2015)1–7.

[12] K. Dermeche, N. Tchouar, S. Belaidi, T. Salah ,Qualitative Structure-Activity Relationships and 2D-QSAR Modeling of TNF- Inhibition by Thalidomide Derivatives, Journal of Bionanoscience 9(2015) 395–400.

[13] S. Belaidi, Z. Almi, D. Bouzidi , Electronic Structure and Physical-Chemistry Properties Relationship for Phenothiazine Derivatives by Quantum Chemical Calculations, Journal of Computational and Theoretical Nanoscience 11(2014) 2481-2488.

[14] S. Khamouli, S. Belaidi, Z. Almi, S. Medjahed, H. Belaidi, Property/Activity Relationships and Drug Likeness for Pyrimidine Derivatives as Serine/Threonine Protein Kinase B Inhibitors, Journal of Bionanoscience 11(2017) 1–9.

[15] D. M. Hawkins The Problem of Overfitting. Journal of chemical information and computer sciences 44 (2004)1-12.

[16] L. Eriksson, J. Jaworska, A. P. Worth, M. T. D. Cronin, M. c. Dowel, P. Gramatica , Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs, Environmental Health Perspectives 111(2003)1361–1375.

[17] T. M. Chapman, N. Bouloc, R. S. Buxton, J. Chugh, K. E. Lougheed, S. A. Osborne, et al, Substituted aminopyrimidine protein kinase B (PknB) inhibitors show activity against Mycobacterium tuberculosis, Bioorganic & Medicinal Chemistry Letters 22(2012)3349-3353 .

[18] M. V. Damre, R. P. Gangwal, G. V. Dhoke, M. Lalit, D.Sharma, K. Khandelwal, A. T. Sangamwar, 3D-QSAR and molecular docking studies of amino-pyrimidine derivatives as PknB inhibitors, Journal of the Taiwan Institute of Chemical Engineers 45(2014)354-364.

[19] HyperChem (Molecular Modeling System) Hypercube (2008) Inc., 1115 NW, 4th Street, Gainesville, FL 32601, USA.

[20] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian 09, Revision A.02, Gaussian, Inc., Wallingford CT (2009).

[21] SPSS software packages, SPSS Inc., 444 North Michigan Avenue, Suite 3000, Chicago, Illinoi, 60611, USA.

[22] A. Golbraikh, A. Tropsha, Beware of q2!, J Mol Graph Model 20(2002) 269–276**.**

[23] A. Tropsha, P. Gramatica, V. K. Gombar,The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models, QSAR & Combinatorial Science 22(2003) 69–77.

[24] K. J. Sanmati, J. Rahul, S. Lokesh, K. Y. Arvind , QSAR Analysis on 3,5-disubstituted-4,5-dihydropyrazole-1-carbothioamides as epidermal growth factor receptor (EGFR) kinase inhibitors, Journal of Chemical and Pharmaceutical Research 4 (2012) 3215-3223.

[25] S. O. Podunavac-Kuzmanovic, D. D. Cvetkovic, D. J. Barna, QSAR Analysis of 2-Amino or 2-Methyl-1-Substituted Benzimidazoles Against Pseudomonas aeruginosa, International Journal of Molecular Sciences 10(2009)1670-1682.

[26] M. Jalali-Heravi , A. Kyani , Use of computer-assisted methods for the modeling of the retention time of a variety of volatile organic compounds: a PCA-MLR-ANN approach, Journal of chemical information and computer sciences 44( 2004) 1328-1335.