Abstract

The quantitative structure-activity relationship (QSAR) of sixty 2-phenylimidazo[1,2-a]pyridines derivatives with anti-Human African Trypanosomiasis (anti-HAT) activity has been studied by using the density functional theory (DFT) and statistical methods. Becke’s three-parameter hybrid method and the Lee-Yang-Parr B3LYP functional employing 6–31G(d) basis set are used to calculate quantum chemical descriptors using Gaussian 03W software, and the five Lipinski’s parameters were calculated using ChemOffice software. In order to obtain robust and reliable QSAR model, the original dataset was randomly divided into training and prediction sets comprising 48 and 12 compounds, respectively. An optimal model for the training set with significant
statistical quality was established. The same model was further applied to predict pEC₅₀ values of the 12 compounds in the test set, further showing that this QSAR model has high predictive ability. It is very interesting to find that the anti-HAT of these compounds appear to be mainly governed by four factors, i.e., the number of H-bond donors, the lowest unoccupied molecular orbital energy, the molecular weight and the octanol/water partition coefficient. Here the possible action mechanism of these compounds was analysed and discussed, in particular, important structural requirements for great anti-HAT activity will be by increasing molecular size and substitute the 2-phenylimidazopyridines derivatives with polar, ionic, stronger accepting electron ability group and heteroatoms attached to one or more hydrogen atoms. Based on this proposed QSAR model, some new compounds with higher anti-HAT activities have been theoretically designed. Such results can offer useful theoretical references for future experimental works.

Keyword: Pharmaceutical chemistry

1. Introduction

Human African Trypanosomiasis (HAT) or African sleeping sickness is one of the infectious diseases grouped under the term Neglected Tropical Diseases, which inflict a devastating effect on the health and economy of nearly 150 countries [1, 2, 3]. In Africa, the number of cases has dropped drastically; however, approximately 3,000 new infections of both East and West African Trypanosomiasis have been reported by the World Health Organization in 2015 [4]. HAT-affected 60 zones covering an area of about 8 million square kilometers between 14° north latitude and 20° south latitude [5]. Most cases of HAT are caused by the protozoans Trypanosoma brucei gambiense and Trypanosoma brucei rhodesiense, which are transmitted to humans through the bite of tsetse flies in rural areas of sub-Saharan Africa.

The disease has two stages: the initial stage is characterized by the spread of the parasite in the blood and the lymphatic system; and in the second stage, the parasite crosses the blood–brain barrier (BBB) in which the parasites spread into the central nervous system [6, 7, 8]. Symptoms of this later stage include sleep disturbance, cognitive dysfunction, coma, and death.

Current treatment for the treatment of HAT includes suramin, pentamidine, melarsoprol, eflornithine, or a combination of nifurtimox and eflornithine (Chemical structures of existing anti-HAT medicines are shown in Fig. 1) [6, 9, 10, 11]. These existing medicines are insufficient, antiquated, toxic, prone to resistance, and require parenteral administration [12, 13, 14, 15, 16, 17]. A new drug, effective for late stage disease that is nontoxic and orally administered, is urgently needed.
In order to open a new way in anti-HAT drug research, a series of sixty 2-phenylimidazopyridine derivatives were synthesized and studied for their anti-HAT activities by Tatipaka et al. [18] to design a better analogue, with rich metabolic stability in liver microsomes, of substituted oxazolopyridine identified as an attractive lead due to good whole-cell activity on Trypanosoma brucei rhodesiense, no cytotoxicity on mammalian cell lines, acceptable exposure in the central nervous system, and satisfactory aqueous solubility.

The main focus of the present study is to develop a QSAR model able to correlate the structural features of the 2-phenylimidazopyridine derivatives with their anti-HAT activities.

In general, the QSAR models are based on the assumption that the activity of a certain chemical compound related to its structure through a certain mathematical algorithm. This relationship can be used in the prediction, interpretation, and assessment of new compounds with desired activities, reducing and rationalizing time, efforts, and cost of synthesis as well as new product development. The basic assumption to drive a QSAR model is presented in the form of a mathematical function associating the chemical properties to the effect (activity). Therefore, the effect is like the function “f” of the chemical properties “x”: $y = f(x)$. To find this algorithm, a number of chemical compounds with known values of the studied effect (y) are considered. For each chemical compound, myriad numbers of parameters (called as chemical descriptors) are calculated. Then, QSAR models are built that provides a quite accurate value, similar to the real experimental value. The final step is to check if the obtained QSAR models are able to predict the activity values for other chemicals not used to build up the model (external validation). Indeed, it is very important to generate a model which worked not only for the chemical
substances used within the training set but also for other similar chemicals. Consequently, the challenge is to define the correct statistical properties of the model [19, 20].

The significance and novelty of findings presented in this work are reflected from the fact that we have used quantum chemistry descriptors which describes electron properties of congeneric structures used in this study, and we have used the five Lipinski’s descriptors to describes compounds could be potential orally administered drugs. The use of density functional theory (DFT) is justified for the reason that some comparative QSAR studies have shown that the descriptors calculated using the DFT method can improve the accuracy of the results and lead to more reliable QSARs [19].

A flow chart for the development of the QSAR model along with the various validation methods used in this work is demonstrated in Fig. 2.

2. Materials and method

2.1. Selection of dataset and generation of molecular descriptors

2.1.1. Data set

In this stage, the data set of the anti-HAT activities of sixty 2-phenylimidazopyridines derivatives were collected from the literature [18, 35]. The molecular structures of the studied molecules with their activity are presented in Fig. 3 and Table 1. All experimental activity values EC$_{50}$ (of compound required to inhibit growth by 50%) were converted to the negative logarithm of EC$_{50}$ (pEC$_{50}$ = -log$_{10}$ (EC$_{50}$)).

Fig. 2. Flow chart of the methodology used in this work.
2.1.2. Molecular descriptors

Electronic and Lipinski’s parameters were calculated using Gaussian 03W and ChemOffice programs, respectively [21, 22], to predict the correlation between these descriptors of the studied molecules with their anti-HAT activities and to develop linear model. Table 2 shows the used descriptors in this study.

2.2. Data set

In this stage, linear QSAR model was developed and evaluated to predict the studied activities of compounds. The descendent multiple linear regression (MLR) analysis available in XLSTAT software [23], based on the elimination of aberrant descriptors (one by one) until a valid model (including the critical probability: p-value < 0.05 for descriptors and for the model), was employed to find a linear model of the activity of interest, which takes the form below:

\[ Y = a_0 + \sum_{i=1}^{n} a_i x_i \]

Where:

- \( Y \): the studied activity, which is, the dependent variable;
- \( a_0 \): the intercept of the equation;
- \( x_i \): the molecular descriptors;
- \( a_i \): the coefficients of those descriptors.

This method is one of the most popular methods of QSAR due to its simplicity in operation, reproducibility and ability to allow easy interpretation of the features used. The important advantage of the linear regression analysis is its transparent nature, therefore, the algorithm is accessible and predictions can be made easily [20].

In order to propose models and to evaluate quantitatively the physicochemical effects of the substituents on the studied activities, we submitted the data matrix constituted obviously from the used variables (descriptors) corresponding to the dataset molecules to a MLR. We use the coefficients R^2, MSE and p-value to select the best regression performance [24]. Where:

1. The R-squared (R^2) also called the coefficient of determination, which is the proportion of variance (%) in the dependent variable that can be explained by the independent variable. Since R^2 value is adopted in various research disciplines, there is no standard guideline to determine the level of predictive acceptance.

Fig. 3. Chemical structures of the studied compounds skeletons.
Table 1. Chemical structures and anti-HAT activities of studied compounds.

| №  | R₁  | R₂   | R₃   | pEC₅₀  |
|----|-----|------|------|--------|
| (A) |     |      |      |        |
| 1  | H   | Cl   |      | 6.658  |
| 2  | H   | F    |      | 6.921  |
| 3  | H   | CH₃  |      | 6.367  |
| 4  | H   | CN   |      | 6.398  |
| 5  | Br  | F    |      | 6.638  |
| 6  | CN  | F    |      | 6.420  |
| 7  | Phenyl | F |      | 7.398  |
| 8 (*) | 4-fluorophenyl | F |      | 7.301  |
| 9  | 3-chlorophenyl | F |      | 7.301  |
| 10 | 4-MeO-phenyl  | F |      | 6.770  |
| 11 (*)| 4-phenylphenyl | F |      | 6.377  |
| (B) |     |      |      |        |
| 12 | 5-methylfuran-2-carbonyl | H | F | 6.699  |
| 13 | 3-methylfuran-2-carbonyl | H | F | 7.000  |
| 14 | 3-furanoyl    | H | F | 6.824  |
| 15 | benzoyl       | H | Cl | 5.149  |
| 16 | oxazole-5-carbonyl | H | F | 5.721  |
| 17 | 2-thiophenyl  | H | Cl | 5.824  |
| 18 (*)| 3-pyridinecarbonyl | H | F | 5.155  |
| 19 | pyrazine-2-carbonyl | H | F | 6.046  |
| 20 (*)| N-methylpyrrole-2-carbonyl | H | F | 5.959  |
| 21 (*)| methylsulfonyl  | H | Cl | 5.215  |
| 22 | 2-furancarbothioyl | H | F | 6.387  |
| 23 | 2-furanoyl    | 2-acetyl | F | 6.301  |
| 24 | 2-furanoyl    | 2-furanoyl | F | 6.921  |
| 25 | benzyl        | Benzyl | F | 5.959  |
| 26 | methylcarbamoyl| H | F | 5.420  |
| 27 | isopropylcarbamoyl | H | Cl | 6.000  |
| 28 | phenylcarbamoyl| H | Cl | 4.921  |
| 29 | dimethylcarbamoyl| H | Cl | 6.398  |
| 30 | 1-pyrrolidinoyl | H | Cl | 7.046  |
| 31 | 1-piperidinoyl | H | Cl | 5.721  |
| (C) |     |      |      |        |
| 32 | 2-furanyl    | H |      | 6.699  |
| 33 (*)| 2-furanyl    | Cl |      | 7.155  |
| 34 | 2-furanyl    | F |      | 6.699  |
| 35 | 2-furanyl    | 5-Cl|      | 5.000  |
| 36 | 2-furanyl    | 7-Cl|      | 6.921  |

(continued on next page)
Henseler et al. [25] and Hair et al [26, 27] proposed a rule of thumb for acceptable $R^2$ with 0.75, 0.50, and 0.25 are described as substantial, moderate and weak respectively. Hence, Zikmund et al. and Moore et al. proposed other rule of thumb for interpreting the strength of a relationship based on its R-squared value (use the absolute value of the R-squared value to make all values positive) [28, 29]:

- If R-squared value $R^2 < 0.3$ this value is generally considered a none or very weak effect size,
- If R-squared value $0.3 < R^2 < 0.5$ this value is generally considered a weak or low effect size,
- If R-squared value $0.5 < R^2 < 0.7$ this value is generally considered a Moderate effect size,

Table 1. (Continued)

| No | $R_1$ | $R_2$ | $R_3$ | $pEC_{50}$ |
|----|-------|-------|-------|------------|
| 37 | N-pyrrolidinyl | Cl | | 7.301 |
| 38 | N-pyrrolidinyl | phenyl | | 8.699 |
| 39 | N-pyrrolidinyl | 3-methoxyphenyl | | 8.000 |
| 40 | N-pyrrolidinyl | 2-methoxyphenyl | | 8.301 |
| 41 | N-pyrrolidinyl | 3-chlorophenyl | | 8.699 |
| 42 | N-pyrrolidinyl | 2-chlorophenyl | | 8.301 |
| 43 | N-pyrrolidinyl | 3-acetylenyl | | 8.523 |
| 44 | N-pyrrolidinyl | 3-methylphenyl | | 8.699 |
| 45 | N-pyrrolidinyl | 3-trifluoromethoxyphenyl | | 7.523 |
| 46 | N-pyrrolidinyl | 3-methyl-4-fluorophenyl | | 7.699 |
| 47 | N-pyrrolidinyl | 3-NH$_2$-phenyl | | 8.000 |
| 48 | N-pyrrolidinyl | 3-furanyl | | 8.398 |
| 49 | N-pyrrolidinyl | 3-thiophenyl | | 8.699 |
| 50 | N-pyrrolidinyl | 2-thiophenyl | | 8.398 |
| 51 | N-pyrrolidinyl | 3-pyridyl | | 8.301 |
| 52 | N-pyrrolidinyl | 5-(2-chloropyridyl) | | 8.000 |
| 53 | N-pyrrolidinyl | 4-(2-chloropyridyl) | | 8.398 |
| 54 | N-pyrrolidinyl | 5-(3-methylpyridyl) | | 8.000 |
| 55 | N-pyrrolidinyl | 5-(2-methoxypyridyl) | | 8.301 |
| 56 | N-pyrrolidinyl | 5-(3-pyridine) | | 7.699 |
| 57 | N-pyrrolidinyl | 5-(3-chloropyrimidine) | | 8.301 |
| 58 | N-pyrrolidinyl | 5-pyrimidinyl | | 7.523 |
| 59 | N-pyrrolidinyl | 5-(2-methoxypyrimidine) | | 7.222 |
| 60 | N-pyrrolidinyl | 5-(2-chloropyrimidine) | | 7.398 |

(*) Test set compounds. $pEC_{50} = -\log_{10}(EC_{50})$. 
Table 2. Values of parameters calculated for the studied compounds.

| N° | E     | µ   | E\textsubscript{HOMO} | E\textsubscript{LUMO} | MW | NHA | NHD | NRB | log P |
|----|-------|-----|------------------------|------------------------|----|-----|-----|-----|-------|
| 1  | -40922.340 | 7.904 | -6.113                | -1.901                 | 340.049 | 6   | 2   | 3   | 3.843 |
| 2  | -31116.982 | 7.495 | -6.062                | -1.826                 | 324.079 | 7   | 2   | 3   | 3.243 |
| 3  | -29486.806 | 6.018 | -5.939                | -1.713                 | 320.104 | 6   | 2   | 3   | 3.989 |
| 4  | -30936.391 | 11.034 | -6.432               | -2.241                 | 331.083 | 6   | 2   | 3   | 3.518 |
| 5  | -101076.674 | 6.975 | -6.153                | -2.043                 | 401.990 | 7   | 2   | 3   | 4.182 |
| 6  | -33626.856 | 7.421 | -6.348                | -2.466                 | 349.074 | 8   | 2   | 4   | 2.921 |
| 7  | -37404.050 | 7.610 | -5.949                | -1.884                 | 400.110 | 7   | 2   | 4   | 5.131 |
| 8  | -40104.179 | 6.960 | -5.983                | -1.936                 | 418.101 | 8   | 2   | 4   | 5.373 |
| 9  | -49909.639 | 8.257 | -6.067                | -2.012                 | 434.071 | 7   | 2   | 4   | 5.854 |
| 10 | -40520.225 | 7.508 | -5.686                | -1.802                 | 430.121 | 8   | 2   | 5   | 5.236 |
| 11 | -43691.129 | 7.708 | -5.839                | -1.910                 | 476.142 | 7   | 2   | 5   | 7.019 |
| 12 | -32186.636 | 5.317 | -6.146                | -1.987                 | 339.102 | 6   | 2   | 3   | 3.796 |
| 13 | -32186.624 | 5.231 | -6.245                | -2.003                 | 337.086 | 6   | 1   | 7   | 2.724 |
| 14 | -31116.792 | 6.390 | -6.030                | -1.870                 | 324.079 | 7   | 2   | 4   | 3.243 |
| 15 | -40982.768 | 7.474 | -6.166                | -1.924                 | 350.070 | 5   | 2   | 3   | 4.667 |
| 16 | -31553.426 | 5.145 | -6.267                | -1.966                 | 327.090 | 8   | 3   | 3   | 2.793 |
| 17 | -49770.221 | 7.034 | -5.890                | -2.125                 | 366.047 | 4   | 2   | 2   | 4.727 |
| 18 | -31613.730 | 5.406 | -6.222                | -1.915                 | 336.090 | 7   | 2   | 3   | 3.314 |
| 19 | -32050.206 | 6.769 | -6.173                | -2.301                 | 336.090 | 8   | 2   | 3   | 2.881 |
| 20 | -31646.188 | 6.852 | -5.952                | -1.909                 | 336.102 | 6   | 1   | 4   | 2.903 |
| 21 | -47607.896 | 5.054 | -6.359                | -2.028                 | 323.013 | 5   | 1   | 3   | 1.935 |
| 22 | -39904.311 | 7.314 | -5.869                | -2.111                 | 340.056 | 6   | 2   | 3   | 3.303 |
| 23 | -35270.067 | 8.211 | -6.247                | -1.968                 | 365.081 | 7   | 0   | 5   | 2.329 |
| 24 | -40426.686 | 5.467 | -6.361                | -1.970                 | 416.084 | 9   | 1   | 5   | 4.774 |
| 25 | -36519.016 | 6.360 | -5.499                | -1.613                 | 410.167 | 6   | 1   | 5   | 6.650 |
| 26 | -27466.414 | 8.042 | -5.960                | -1.761                 | 286.087 | 5   | 2   | 4   | 1.896 |
| 27 | -39411.369 | 8.621 | -5.993                | -1.822                 | 330.088 | 4   | 2   | 5   | 3.053 |
| 28 | -42489.155 | 1.786 | -5.891                | -1.916                 | 365.081 | 5   | 3   | 4   | 4.707 |
| 29 | -38341.359 | 8.507 | -5.957                | -1.812                 | 316.073 | 4   | 1   | 4   | 2.091 |
| 30 | -40448.111 | 9.131 | -5.884                | -1.777                 | 343.097 | 5   | 2   | 3   | 3.828 |
| 31 | -41517.874 | 8.542 | -5.927                | -1.804                 | 357.112 | 5   | 2   | 3   | 4.387 |
| 32 | -30576.266 | 3.031 | -6.117                | -1.602                 | 323.095 | 7   | 3   | 3   | 3.625 |
| 33 | -43081.767 | 4.761 | -6.292                | -1.885                 | 357.056 | 7   | 3   | 3   | 4.483 |
| 34 | -33276.260 | 4.502 | -6.216                | -1.786                 | 341.086 | 8   | 3   | 3   | 4.002 |
| 35 | -43081.927 | 4.900 | -6.281                | -1.815                 | 357.056 | 7   | 3   | 3   | 4.483 |
| 36 | -43081.878 | 1.860 | -6.339                | -1.780                 | 357.056 | 7   | 3   | 3   | 4.483 |
| 37 | -42608.017 | 4.941 | -6.020                | -1.827                 | 360.103 | 6   | 3   | 3   | 3.179 |
| 38 | -36389.529 | 4.940 | -5.774                | -1.680                 | 402.174 | 6   | 3   | 4   | 4.209 |

(continued on next page)
If R-squared value $R^2 > 0.7$ this value is generally considered strong effect size.

2. The mean square error (MSE): measure the average squared difference between the predicted and experimental activities values.

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_{i}^{pred} - y_{i}^{obs})$$

N: The number of data points  
$y_{i}^{pred}$: The predicted (calculated by the model) value for data point i.  
$y_{i}^{obs}$: The actual (observed) value for data point i.

The MSE is always strictly positive, and a good model will be with MSE values closer to zero, i.e. the good model will be with minimize the sum of the squared difference between the true and estimated values [30].

3. $P$-value: the significance level, which gives an indication of the probability that a QSAR is a significant occurrence.

| N  | E     | $\mu$ | $E_{\text{HOMO}}$ | $E_{\text{LUMO}}$ | MW   | NHA | NHD | NRB | log P |
|----|-------|-------|-------------------|-------------------|------|-----|-----|-----|-------|
| 39 | -39505.695 | 4.609 | -5.694 | -1.673 | 432.184 | 7   | 3   | 5   | 4.236 |
| 40 | -39505.605 | 4.478 | -5.577 | -1.544 | 432.184 | 7   | 3   | 5   | 3.676 |
| 41 | -48895.128 | 5.536 | -5.913 | -1.803 | 436.135 | 6   | 3   | 4   | 4.951 |
| 42 | -48895.018 | 5.269 | -5.881 | -1.704 | 436.135 | 6   | 3   | 4   | 4.701 |
| 43 | -40543.039 | 6.462 | -5.862 | -1.776 | 444.184 | 7   | 3   | 5   | 3.749 |
| 44 | -37459.374 | 4.947 | -5.736 | -1.655 | 416.189 | 6   | 3   | 4   | 4.708 |
| 45 | -47606.991 | 5.648 | -5.912 | -1.805 | 486.156 | 10  | 3   | 6   | 5.345 |
| 46 | -40159.542 | 4.670 | -5.759 | -1.699 | 434.180 | 7   | 3   | 4   | 4.880 |
| 47 | -37895.697 | 6.101 | -5.482 | -1.614 | 417.184 | 7   | 4   | 4   | 3.162 |
| 48 | -36328.934 | 4.630 | -5.691 | -1.680 | 392.153 | 7   | 3   | 4   | 3.385 |
| 49 | -45117.304 | 5.511 | -5.726 | -1.688 | 408.130 | 6   | 3   | 4   | 5.213 |
| 50 | -45117.314 | 5.785 | -5.634 | -1.763 | 408.130 | 6   | 3   | 4   | 5.423 |
| 51 | -36825.916 | 3.244 | -5.931 | -1.806 | 403.169 | 7   | 3   | 4   | 4.245 |
| 52 | -49331.575 | 4.023 | -6.029 | -1.933 | 437.130 | 7   | 3   | 4   | 5.010 |
| 53 | -49331.591 | 4.764 | -6.111 | -2.046 | 437.130 | 7   | 3   | 4   | 5.010 |
| 54 | -37895.779 | 2.889 | -5.899 | -1.776 | 417.184 | 7   | 3   | 4   | 4.833 |
| 55 | -39942.234 | 2.459 | -5.734 | -1.714 | 433.179 | 8   | 3   | 5   | 5.154 |
| 56 | -42577.963 | 2.351 | -5.259 | -1.654 | 472.227 | 8   | 3   | 5   | 5.112 |
| 57 | -49331.478 | 4.961 | -6.042 | -1.940 | 437.130 | 7   | 3   | 4   | 5.010 |
| 58 | -37262.327 | 5.486 | -6.061 | -1.960 | 404.164 | 8   | 3   | 4   | 3.338 |
| 59 | -40378.895 | 3.634 | -5.918 | -1.830 | 434.175 | 9   | 3   | 5   | 4.373 |
| 60 | -49767.937 | 6.555 | -6.138 | -2.084 | 438.125 | 8   | 3   | 4   | 4.065 |
In order to assess the significance of the models and its accurate prediction ability for new compounds:

4. The variance inflation factor VIF [31] to detect the absence of the multicollinearity between descriptors was used; models with descriptors correlated with each other are not significant. The VIF was defined as $1/(1-R^2)$, where $R$ is the coefficient of correlation between one descriptor and all other descriptors in model. A VIF value greater than 5.0 indicates that the model is unstable; a value between 1.0 and 4.0 indicates that the model is acceptable [32].

5. In addition, an internal validation procedure (leave-one-out cross validation) was employed, in which one compound is removed and the rebuilt model with the remaining molecules is used to predict the response of the eliminated compound. This one is then returned and a second is removed, and the cycle is repeated, and so on until all compounds have been removed one by one, and an overall correlation coefficient $R_{cv}$ is computed [33]. A model is considered acceptable when the value of $R_{cv}^2$ exceeds 0.6 [27, 32].

6. After the model is built, an external prediction is necessary. This one remains the only way to determine both the generalizability of QSAR model for new chemicals and the true predictive power of the models. In this external validation, the obtained model was used to predict the activities of a test set comprising compounds that are similar to though not used in the training set. This is usually performed by splitting a data set into a training set a test set, typically in a 1:5 ratio [34].

7. A model is valid only within its training domain and new molecules must be considered as belonging to the domain before the model is applied (OECD Principle 3 [35]). Without applicability domain (AD), each model can predict the activity of any compound, even with a completely different structure from those included in the study. Therefore, the AD is a tool to find out compounds that are outside of the built QSAR model and it detects outliers present in the training set compounds. There are several methods for defining the applicability domain (AD) of QSAR models [36], but the most common one is determining the leverage values $h_i$ ($h_i = x_i^T (X^T X)^{-1} x_i$) for each compound, where: $x_i$: the descriptor row-vector of query compound, $X$: the $n \times (k-1)$ matrix of $k$ model descriptor values for $n$ training set compounds and the superscript $t$ refers to the transpose of matrix/vector [37]. In this study, we use the Williams plot; in this plot, the applicability domain is established inside a squared area within standard deviation ± x (in this study $x = 2.5$; “three sigma rule” [38]) and a leverage threshold $h^*$ ($h^* = 2.5 \times (k+1)/n$) [39]. Where: $n$ is the number of training set compounds, $k$ is the number of model descriptors. The leverage ($h$) greater than the warning leverage ($h^*$) suggested that the compound was very influential on the model [40].

8. Further, the y-randomization approach was performed to ensure the robustness of a predictive model. Often, it is used along with the cross-validation. It consists of
repeating the calculation procedure with randomized activities and subsequent probability assessment of the resultant statistics. The dependent variable vector is randomly shuffled and a new predictive model is developed using the original independent variable matrix. The new predictive models (after several repetitions) are expected to have low $R^2$ and $R^2_{cv}$ values. If the opposite happens, then an acceptable model cannot be obtained for the specific modelling method and data [41]. Another parameter, $cR^2_p$ is also calculated which should be more than 0.5 for passing this test: 

$$cR^2_p = R^* \sqrt{(R^2 - \text{Average } Rr)^2}.$$ 

Where: Average $Rr$ = average ‘R’ of random models [42].

### 3. Results and discussions

#### 3.1. Molecular descriptors

From the results of the density functional theory DFT (B3LYP/6-31G (d)) calculations, following quantum chemistry descriptors were obtained for building the model: total energy $E$, dipole moment $\mu$, highest occupied molecular orbital energy $E_{HOMO}$ and lowest unoccupied molecular orbital energy $E_{LUMO}$.

The five Lipinski’s parameters calculated are: molecular weight $MW$, number of H-bond acceptors $NHA$, number of H-bond donors $NHD$, number of rotatable bonds $NRB$ and octanol/water partition coefficient $\log \text{P}$ (Table 2).

#### 3.2. Multiple linear regression (MLR)

The QSAR analysis was performed using calculated molecular descriptors and the experimental values of the anti-HAT activities for the forty-eight 2-phenylimidazopyridines derivatives (effect). The established MLR model is represented by the following equation along with the values of the statistical parameters:

$$pEC_{50} = 3.036 + 1.196 \text{E}_{LUMO} + 0.019 \text{MW} + 0.301 \text{NHD} - 0.398 \log \text{P}$$  \hspace{1cm} (1)

$$R^2 = 0.598; \hspace{0.1cm} R = 0.773; \hspace{0.1cm} \text{MSE} = 0.508; \hspace{0.1cm} \text{p-value} < 10^{-4}$$

The values of calculated activities from Eq. (1) have been presented in Table 3 and the correlations of calculated and observed activities values are illustrated in Fig. 4.

The $p$-value is lower than 0.0001, it means that we would be taking a lower than 0.01% risk in assuming that the null hypothesis (no effect of the explanatory variables) is wrong. Equation has statistically significance. Therefore, we can conclude with confidence that the selected variables do bring a significant amount of information.

The value of $R^2$ and the MSE indicate that the proposed model is predictive and reliable.
The VIF values of all four descriptors in MLR model are smaller than 4.0 (VIF = 1.173, 2.170, 1.270 and 2.094 for ELUMO, MW, NHD and log P, respectively) indicating that there is no collinearity among the selected descriptors and the resulting model has good stability.

The obtained model was validated internally by the leave-one-out cross validation technique, the cross-validation coefficient $R^2_{cv}$ for the model was determined based on the predictive ability of the model. The value of $R^2_{cv}$ is higher than 0.5 ($R^2_{cv} = 0.509$), it indicates the better predictively of the model.

True predictive power of this model is to test their ability to predict perfectly the pEC$_{50}$ of compounds from an external test set (compounds that were not used for the developed model), the pEC$_{50}$ of the remained set of 12 compounds are deduced from the quantitative model proposed with the compounds used in training set by

| N  | pEC$_{50}$ | MLR | Residues | N  | pEC$_{50}$ | MLR | Residues |
|----|-----------|-----|----------|----|-----------|-----|----------|
| 1  | 6.658     | 6.250 | 0.408    | 30 | 7.046     | 6.463 | 0.584    |
| 2  | 6.921     | 6.278 | 0.643    | 31 | 5.721     | 6.472 | -0.751   |
| 3  | 6.367     | 6.041 | 0.326    | 32 | 6.699     | 6.676 | 0.024    |
| 4  | 6.398     | 5.804 | 0.595    | 34 | 6.699     | 6.645 | 0.054    |
| 5  | 6.638     | 7.115 | -0.477   | 35 | 5.000     | 6.720 | -1.720   |
| 6  | 6.420     | 6.112 | 0.308    | 36 | 6.921     | 6.762 | 0.159    |
| 7  | 7.398     | 6.891 | 0.507    | 37 | 7.301     | 7.283 | 0.019    |
| 9  | 7.301     | 7.091 | 0.210    | 38 | 8.699     | 7.842 | 0.857    |
| 10 | 6.770     | 7.514 | -0.744   | 39 | 8.000     | 8.406 | -0.406   |
| 12 | 6.699     | 6.148 | 0.551    | 41 | 8.699     | 8.040 | 0.660    |
| 13 | 7.000     | 6.217 | 0.783    | 42 | 8.301     | 8.258 | 0.043    |
| 14 | 6.824     | 6.225 | 0.599    | 43 | 8.523     | 8.703 | -0.180   |
| 15 | 5.149     | 6.084 | -0.935   | 44 | 8.699     | 7.938 | 0.761    |
| 16 | 5.721     | 6.647 | -0.926   | 48 | 8.398     | 7.981 | 0.417    |
| 17 | 5.824     | 6.121 | -0.297   | 49 | 8.699     | 7.545 | 1.154    |
| 19 | 6.046     | 6.081 | -0.035   | 50 | 8.398     | 7.371 | 1.027    |
| 22 | 6.387     | 6.214 | 0.173    | 51 | 8.301     | 7.695 | 0.606    |
| 23 | 6.301     | 6.643 | -0.342   | 53 | 8.398     | 7.744 | 0.654    |
| 24 | 6.921     | 6.968 | -0.047   | 54 | 8.000     | 7.762 | 0.238    |
| 25 | 5.959     | 6.499 | -0.540   | 55 | 8.301     | 8.010 | 0.291    |
| 26 | 5.420     | 6.174 | -0.754   | 56 | 7.699     | 8.835 | -1.136   |
| 27 | 6.000     | 6.472 | -0.472   | 57 | 8.301     | 7.872 | 0.429    |
| 28 | 4.921     | 6.661 | -1.740   | 59 | 7.222     | 8.201 | -0.979   |
| 29 | 6.398     | 6.301 | 0.097    | 60 | 7.398     | 8.094 | -0.696   |
MLR. This model will be able to predict the activities of test set molecules in agreement with the experimentally determined value. The observed and calculated pEC50 values are given in Table 4. The predictive capacity of the models that was judged, the higher value of $R^2_{\text{test}}$ ($R^2_{\text{test}} = 0.700$) indicate the improved predictively of the model.

In the Eq. (1), the number of H-bond donors NHD, the lowest unoccupied molecular orbital energy $E_{\text{LUMO}}$ and the molecular weight MW influence positively the

![Correlations of observed and predicted activities (training set in blue and test set in red) values calculated using MLR models.](image)

**Fig. 4.** Correlations of observed and predicted activities (training set in blue and test set in red) values calculated using MLR models.

**Table 4.** Chemical descriptors, observed and MLR predicted activities for the test set.

| Test Set | pEC50 | E_HOMO | MW  | NHD | logP | Obs.  | MLR  |
|----------|-------|--------|-----|-----|------|-------|------|
| 8        | -1.936| 418.101| 2   | 5.373| 7.301| 7.072 |
| 11       | -1.910| 476.142| 2   | 7.019| 6.377| 7.543 |
| 18       | -1.915| 335.095| 2   | 3.314| 5.155| 6.351 |
| 20       | -1.909| 336.102| 1   | 2.903| 5.959| 6.239 |
| 21       | -2.028| 323.013| 1   | 1.935| 5.215| 6.235 |
| 33       | -1.885| 357.056| 3   | 4.832| 7.155| 6.636 |
| 40       | -1.544| 432.184| 3   | 3.676| 8.301| 8.783 |
| 45       | -1.805| 486.156| 3   | 5.345| 7.523| 8.824 |
| 46       | -1.699| 434.180| 3   | 4.880| 7.699| 8.156 |
| 47       | -1.614| 417.184| 4   | 3.162| 8.000| 8.922 |
| 52       | -1.933| 437.130| 3   | 5.010| 8.000| 7.880 |
| 58       | -1.960| 404.164| 3   | 3.338| 7.523| 7.892 |
activities and the octanol/water partition coefficient log P influence negatively the activities.

By definition:

- Molecular weight MW is the sum of the masses of all the atoms in the molecular formula of the molecule. This descriptor has been used as a descriptor in systems such as transport studies where diffusion is the mode of operation. It is an important variable in QSAR studies pertaining to cross resistance of various drugs in multidrug resistant cell lines [43, 44]. For orally delivered drugs, the molecular weight must be less than or equal to 500 Daltons [45].

- Partition coefficient octanol-water (Log P) is the ratio of concentrations of a substance in a mixture of two solvents, octanol and water. Both solvents are immiscible and therefore form two phases [46].

\[ \text{LogP} = \log \frac{[\text{Octanol}]}{[\text{Water}]} \]

The LogP is the most useful parameter for the characterization of hydrophobicity (and polarity) of compounds [47]. It is an important variable in QSAR studies because the distribution of chemicals between fatty and aqueous phases of a biological system could totally account for the variation in activities [48]. For orally delivered drugs, the partition coefficient octanol-water must be less than 5 [45].

- The number of H-bond Donors NHD is a crucial descriptor in the description of diverse processes occurring in condensed media such as dissolving, partitioning, solubilisation, etc. Drug action and bioavailability critically depends on aqueous solubility, blood-tissue distribution, and specifically on hydrogen binding to receptor active sites and transport proteins [49]. For orally delivered drugs, the hydrogen bond donors must be less than 5 [45].

- The lowest unoccupied molecular orbital energy, \( E_{\text{LUMO}} \): HOMO and LUMO refer to highest occupied molecular orbital and lowest unoccupied molecular orbital. According to the frontier orbital theory, the nucleophilic attack occurs by electron flow from the HOMO of the nucleophile into the LUMO of the electrophile. In stable molecules, occupied electrons always reside into orbitals with negative energies and unoccupied orbitals have positive energies. The energies of HOMO and LUMO are related to the reactivity of the molecule: molecules with electrons at accessible (near-zero) HOMO levels tend to be good nucleophiles because it does not cost much to donate these electrons toward making a new bond. Similarly, molecules with lower LUMO energies tend to be good electrophiles because it does not cost much to place an electron into such an orbital [50, 51].

Comparing the importance of each descriptor on pEC50 of 2-phenylimidazopyridines, one must know the standardized coefficient or the t-test values of them in the model.
The bigger the absolute value of the t-test value is, the greater the influence of the descriptor is. The t-test values for our model descriptors are 1.928, 5.714, 1.883 and -2.639 for $E_{\text{LUMO}}$, MW, NHD, and logP, respectively. This means that the t-test value of logP is larger than that of other three descriptors, which indicate that in this model, the influence of MW on activity is stronger than that of the others.

Consequently, if we want to increase the value of the activity, we will:

- Decrease the logP (with negative sign in the model) value, for which we must substitute the 2-phenylimidazopyridines derivatives for hydrophilic ("water-loving") substituents. This means that a substitution with a polar and ionic group (such as: -OH, -COOH, -NH$_2$) may lead to high activity values.

- Increase the $E_{\text{LUMO}}$ value, for which we suggest the substitution of the 2-phenylimidazopyridines derivatives with a stronger accepting electron ability group, positively charged or neutral species having vacant orbitals that are attracted to an electron rich centre (such as ROX, BH, -NO$_2$).

- Increase the NHD (with positive sign in the model) value, for which substituting the 2-phenylimidazopyridines derivatives by heteroatom attached to one or more hydrogen atoms.

- Increase the MW value, for which increasing the molecular size.

In the conclusion, these results illustrate that to increase the anti-HAT, we will increase the molecular size and substitute the 2-phenylimidazopyridines derivatives with polar, ionic, stronger accepting electron ability group and heteroatoms attached to one or more hydrogen atoms.

This study is in agreement with the conclusions of a previous QSAR studies [52] which revealed the importance the presence of five membered rings, especially the pyrrolidine ring, is beneficial for the HAT activity of the present series of molecules; and the interesting pattern of H-bond donor/acceptor nitrogen atoms in attaining various tautomeric forms, thereby, providing additional flexibility to the molecules to acquire bioactive tautomeric form(s) while interacting with the target receptor.

Further, before performing the external validation of a model, it is very important to check for the presence of systematic error that violates the basic assumptions of the least squares regression model. If high systematic error (bias) is present in the model, then such model should be discarded and performing any external validation test is of no use on such biased model. Xternal Validation Plus is a tool that checks the presence of systematic errors in the model and further computes all the required external validation parameters, while judging the performance of actual prediction quality of a QSAR model based on recently proposed MAE-based criteria [53];
Table 5. Output file summarize the information including all the external validation parameters that are required to judge the performance of prediction quality of the MLR model.

| User Input File Info. | File Name       | Sample_TestSet.xlsx |
|-----------------------|-----------------|---------------------|
| Model biasness test   | nPE/nNE         | 0.3333              |
|                       | nNE/nPE         | 3.0000              |
|                       | [MPE/MNE]       | 0.3620              |
|                       | [MNE/MPE]       | 2.7623              |
|                       | AAE - [AE]      | 0.1447              |
|                       | R2 (Residuals; serial correlation) | 0.0421 |
|                       | R2 (Residuals and Yobs values) | 0.1383 |
|                       | R²Test (100% data) | 0.6997 |
|                       | R₀²Test (100% data) | 0.6978 |
|                       | R₀²Test (100% data) | 0.6123 |
| Classical Metrics     | Q2F1 (100% data) | 0.4622              |
|                       | Q2F2 (100% data) | 0.4291              |
|                       | Scaled Avg.Rm² (100% data) | 0.5094 |
|                       | Scaled DeltaRm² (100% data) | 0.2532 |
|                       | CCC(100% data)  | 0.7351              |
|                       | R²Test (95% data) | 0.7330              |
|                       | R₀²Test (95% data) | 0.7321 |
|                       | R₀²Test (95% data) | 0.4763 |
|                       | Q2F1 (95% data)  | 0.5674              |
|                       | Q2F2 (95% data)  | 0.5506              |
|                       | ScaledAvgRm² (95% data) | 0.5028 |
|                       | Scaled DeltaRm² (95% data) | 0.2536 |
|                       | CCC(95% data)   | 0.7703              |
| Error-based metrics (for 100% data) | RMSEP (100% data) | 0.7834 |
|                       | SD (100% data)  | 0.4210              |
|                       | SE (100% data)  | 0.1215              |
|                       | MAE (100% data) | 0.6718              |
|                       | RMSEP (95% data) | 0.7181              |
|                       | SD (95% data)   | 0.3896              |
|                       | SE (95% data)   | 0.1175              |
|                       | MAE (95% data)  | 0.6145              |
|                       | MAE+3*SD (95% data) | 1.7834 |
|                       | NCompTest       | 12.0000             |

Number of test set compounds. Train range 7.8195 (continued on next page)
Xternal Validation Plus indicates the absence of systematic errors in the model and a moderate performance of prediction quality of a QSAR model based on proposed MAE-based criteria (Table 5).

In the next step, all calculations were repeated with randomized activities of the training set compounds as well to evaluate model robustness (y-randomization test). In the present case, 100 random trials were run for the MLR model. None of the random trials could match the original model (Table 6). The standalone QSAR-tools (“Programs”) available online at http://dtclab.webs.com/software-tools and http://teqip.jdvu.ac.in/QSAR_Tools/ (“Websites”) was employed in the y-randomization.

The average value of R, R² and $R^2_{CV}$ are 0.282, 0.087 and -0.155 respectively, the $cR^2_{p}$ value equal a 0.557 (more than 0.5), and all the new QSAR models having significantly low R² and $R^2_{CV}$ values for the 100 trials, which confirm that the developed QSAR models are robust.

The applicability domain (AD) of the MLR models was evaluated by leverage analysis expressed as Williams plot (Fig. 5), in which the standardized residuals and the
Table 6. Y-randomization test results for MLR models.

| Random | R   | R²  | R²CV | Random | R   | R²  | R²CV | Random | R   | R²  | R²CV | Random | R   | R²  | R²CV |
|--------|-----|-----|------|--------|-----|-----|------|--------|-----|-----|------|--------|-----|-----|------|
| 1      | 0.119 | 0.014 | -0.319 | 26 | 0.141 | 0.020 | -0.246 | 51 | 0.212 | 0.045 | -0.230 | 76 | 0.219 | 0.048 | -0.185 |
| 2      | 0.382 | 0.146 | -0.049 | 27 | 0.398 | 0.158 | -0.122 | 52 | 0.278 | 0.077 | -0.113 | 77 | 0.202 | 0.041 | -0.222 |
| 3      | 0.419 | 0.176 | 0.013 | 28 | 0.242 | 0.059 | -0.270 | 53 | 0.349 | 0.122 | -0.073 | 78 | 0.249 | 0.062 | -0.248 |
| 4      | 0.212 | 0.045 | -0.159 | 29 | 0.388 | 0.150 | -0.072 | 54 | 0.251 | 0.063 | -0.235 | 79 | 0.341 | 0.116 | -0.112 |
| 5      | 0.196 | 0.039 | -0.177 | 30 | 0.133 | 0.018 | -0.224 | 55 | 0.354 | 0.125 | -0.136 | 80 | 0.154 | 0.024 | -0.192 |
| 6      | 0.368 | 0.135 | -0.148 | 31 | 0.273 | 0.074 | -0.165 | 56 | 0.282 | 0.079 | -0.096 | 81 | 0.314 | 0.099 | -0.178 |
| 7      | 0.459 | 0.210 | -0.040 | 32 | 0.317 | 0.100 | -0.130 | 57 | 0.239 | 0.057 | -0.251 | 82 | 0.166 | 0.028 | -0.184 |
| 8      | 0.390 | 0.152 | -0.022 | 33 | 0.232 | 0.054 | -0.169 | 58 | 0.270 | 0.073 | -0.144 | 83 | 0.302 | 0.091 | -0.121 |
| 9      | 0.319 | 0.102 | -0.139 | 34 | 0.282 | 0.079 | -0.175 | 59 | 0.300 | 0.090 | -0.151 | 84 | 0.347 | 0.120 | -0.224 |
| 10     | 0.323 | 0.104 | -0.177 | 35 | 0.169 | 0.029 | -0.194 | 60 | 0.296 | 0.088 | -0.201 | 85 | 0.377 | 0.142 | -0.066 |
| 11     | 0.176 | 0.031 | -0.235 | 36 | 0.166 | 0.028 | -0.195 | 61 | 0.283 | 0.080 | -0.107 | 86 | 0.256 | 0.066 | -0.168 |
| 12     | 0.293 | 0.086 | -0.145 | 37 | 0.174 | 0.030 | -0.253 | 62 | 0.336 | 0.113 | -0.086 | 87 | 0.318 | 0.101 | -0.225 |
| 13     | 0.488 | 0.238 | 0.104 | 38 | 0.403 | 0.162 | -0.072 | 63 | 0.404 | 0.164 | -0.007 | 88 | 0.398 | 0.158 | -0.057 |
| 14     | 0.377 | 0.142 | -0.052 | 39 | 0.243 | 0.059 | -0.145 | 64 | 0.442 | 0.195 | 0.003 | 89 | 0.253 | 0.064 | -0.139 |
| 15     | 0.317 | 0.101 | -0.119 | 40 | 0.458 | 0.210 | -0.025 | 65 | 0.317 | 0.101 | -0.109 | 90 | 0.458 | 0.210 | 0.004 |
| 16     | 0.395 | 0.156 | -0.086 | 41 | 0.261 | 0.068 | -0.197 | 66 | 0.348 | 0.121 | -0.160 | 91 | 0.396 | 0.157 | -0.068 |
| 17     | 0.245 | 0.060 | -0.268 | 42 | 0.166 | 0.028 | -0.230 | 67 | 0.164 | 0.027 | -0.228 | 92 | 0.430 | 0.185 | -0.070 |
| 18     | 0.264 | 0.070 | -0.150 | 43 | 0.231 | 0.053 | -0.193 | 68 | 0.263 | 0.069 | -0.299 | 93 | 0.256 | 0.065 | -0.150 |
| 19     | 0.246 | 0.060 | -0.154 | 44 | 0.201 | 0.041 | -0.235 | 69 | 0.210 | 0.044 | -0.202 | 94 | 0.279 | 0.078 | -0.224 |
| 20     | 0.305 | 0.093 | -0.138 | 45 | 0.271 | 0.073 | -0.122 | 70 | 0.285 | 0.081 | -0.104 | 95 | 0.194 | 0.038 | -0.191 |
| 21     | 0.154 | 0.024 | -0.179 | 46 | 0.225 | 0.051 | -0.238 | 71 | 0.190 | 0.036 | -0.287 | 96 | 0.260 | 0.068 | -0.122 |
| 22     | 0.269 | 0.073 | -0.257 | 47 | 0.336 | 0.113 | -0.075 | 72 | 0.349 | 0.122 | -0.080 | 97 | 0.204 | 0.042 | -0.268 |
| 23     | 0.151 | 0.023 | -0.225 | 48 | 0.373 | 0.139 | -0.071 | 73 | 0.285 | 0.081 | -0.142 | 98 | 0.324 | 0.105 | -0.182 |
| 24     | 0.286 | 0.082 | -0.149 | 49 | 0.261 | 0.068 | -0.171 | 74 | 0.210 | 0.044 | -0.192 | 99 | 0.279 | 0.078 | -0.261 |
| 25     | 0.142 | 0.020 | -0.313 | 50 | 0.364 | 0.132 | -0.088 | 75 | 0.219 | 0.048 | -0.152 | 100 | 0.120 | 0.014 | -0.217 |

Average

| R   | R²  | R²CV | R²CV |
|-----|-----|------|------|
| 0.282 | 0.087 | -0.155 | 0.557 |
leverage threshold values ($h^* = 0.260$) were plotted. Any new value of predicted $p\text{EC}_{50}$ data must be considered reliable only for those compounds that fall within this AD on which the model was constructed.

From the Fig. 5, it is obvious that all the compounds have a standard deviation into the $\pm x$ interval ($x = 2.5$) and there is two responses outliers both in training set and no response outside in test set. These outliers (compounds 23 and 25) have a higher leverage which is greater than $h^*$ value of 0.260. These erroneous predictions could probably be attributed to the structural of these outsides (Fig. 6); maybe the selected

![Leverages h](image)

**Fig. 5.** Williams plot of standardized residual versus leverage for the MLR model (With: $h^* = 0.260$ and residual limits $= \pm 2.5$); Train samples in black colour and test samples in red colour).

**Fig. 6.** Chemical structures of the outsides compounds.
descriptors do not pay much attention to these substructures or their mechanism of action may be different. The predictions of these two compounds are extrapolations of the model, but fortunately they are all ‘‘good leverage’’ chemicals.

The results obtained by MLR are very sufficient to conclude the performance of the models. Consequently, we can design new compounds with improved values of activity than the studied compounds using this model.

Taking into account the above results, we added suitable substitutions and then we moved to calculate their activities using the proposed models Eq. (1). Therefore, the suggested model will reduce the time and cost of synthesis as well as the determination of the anti-HAT activity for the 2-phenylimidazopyridines derivatives.

According to the above discussions, the MLR model could be applied to other 2-phenylimidazopyridines derivatives accordingly to Table 1 and could add further knowledge in the improvement of new way in anti-HAT drug research. If we develop a new compound with better values than the existing ones, it may give rise to the development of more active compounds than those currently in use.

In this way, we carried out structural modification starting from compounds having the highest pEC$_{50}$ values as template (38, 41, 44, and 49). The structures of the designed compounds and their parameter values calculated by the same methods, as well as the pEC$_{50}$ values theoretically predicted by the MLR model are listed in Table 7.

**Table 7.** Values of descriptors, calculated anti-HAT activity pEC$_{50}$ and leverages (h) for the new designed compounds (derivatives of the skeleton (C) of the Fig. 1).

| Designed compounds | E$_{LUMO}$ | MW     | NHD | logP | pEC$_{50}$ | leverage |
|--------------------|-----------|--------|-----|------|-----------|----------|
| X$_1$              | R$_1$=COOH; R$_2$=COOH | 421.095 | 5   | 2.773 | 8.837     | 0.436    |
| X$_2$              | R$_1$=NO$_2$; R$_2$= NO$_2$ | 427.117 | 7   | 0.745 | 9.125     | 1.596    |
| X$_3$              | R$_1$=NH$_2$; R$_2$= p-PhNH$_2$ | 363.138 | 5   | 1.796 | 8.669     | 0.588    |
| X$_4$              | R$_1$=N-pyrrolidinyl; R$_2$= m-PhNO$_2$ | 446.163 | 4   | 4.048 | 8.919     | 0.170    |
| X$_5$              | R$_1$=N-pyrrolidinyl; R$_2$= p-PhNO$_2$ | 446.163 | 4   | 4.048 | 8.735     | 0.145    |
| X$_6$              | R$_1$=N-pyrrolidinyl; R$_2$= p-PhOH | 418.168 | 4   | 3.733 | 8.698     | 0.191    |
| X$_7$              | R$_1$=N-pyrrolidinyl; R$_2$= m,p-Ph(OH)$_2$ | 434.163 | 3   | 3.327 | 8.834     | 0.208    |
| X$_8$              | R$_1$= N-pyrrolidinyl; R$_2$=3-thio-2,4-dihydroxyphenyl | 440.120 | 5   | 3.926 | 9.498     | 0.395    |
| X$_9$              | R$_1$= N-pyrrolidinyl; R$_2$=3-methyl-4-hydroxyphenyl | 432.184 | 4   | 4.182 | 8.817     | 0.176    |
| X$_{10}$           | R$_1$= N-pyrrolidinyl; R$_2$=3-methyl-6-hydroxyphenyl | 432.184 | 4   | 3.932 | 8.984     | 0.214    |
| X$_{11}$           | R$_1$= N-pyrrolidinyl; R$_2$=3-methyl-5-hydroxyphenyl | 432.184 | 4   | 4.232 | 8.753     | 0.164    |
4. Conclusion

The results of the QSAR analysis suggest that derivatives of 2-phenylimidazopyridines with the following structural feature may exhibit great anti-HAT activity by increasing molecular size and substitute the 2-phenylimidazopyridines derivatives with polar, ionic, stronger accepting electron ability group and heteroatoms attached to one or more hydrogen atoms.

According to developed model, the most important findings of this research are that we have designed and suggest some new compounds with possible great activities. Consequently, the proposed models can be used in anti-HAT drug research for the 2-phenylimidazopyridines derivatives. These results encourage the collaboration between theoretical researchers and pharmacologists, academic or industrial, because the last ones many times are groping new drugs.

Declarations

Author contribution statement

Samir Chtita: Analyzed and interpreted the data; Wrote the paper.

Mounir Ghamali, Abdellah Ousaa, Assia Belhassan, Abdelali Idrissi Taourati: Analyzed and interpreted the data.

Adnane Aouidate: Conceived and designed the experiments.

Vijay Hariram Masand: Contributed reagents, materials, analysis tools or data; Wrote the paper.

Mohammed Bouachrine: Contributed reagents, materials, analysis tools or data.

Tahar Lakhlif: Conceived and designed the experiments.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interest statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.
Acknowledgements

We are grateful to the “Association Marocaine des Chimistes Théoriciens” (AMCT) for its pertinent help concerning the programs.

References

[1] P.J. Hotez, A. Fenwick, L. Savioli, D.H. Molyneux, Rescuing the bottom billion through control of neglected tropical diseases, Lancet 373 (2009) 1570–1575.

[2] K. Chibale, New developments in antiinfectives research for tropical infectious diseases, Bioorg. Med. Chem. 23 (16) (2015) 5085–5086.

[3] M. Njoroge, N.M. Njuguna, P. Mutai, D.S.B. Ongarora, P.W. Smith, K. Chibale, Recent approaches to chemical discovery and development against malaria and the neglected tropical diseases human African trypanosomiasis and schistosomiasis, Chem. Rev. 114 (2014) 11138–11163.

[4] www.who.int/mediacentre/factsheets/fs259/en/. (Accessed 27 February 2019).

[5] D. Steverding, The history of African trypanosomiasis, Parasites Vectors 3—3 (2008) 1–8.

[6] R. Brun, J. Blum, F. Chappuis, C. Burri, Human African trypanosomiasis, Lancet 375 (2010) 148–159.

[7] R. Brun, R. Don, R.T. Jacobs, M.Z. Wang, M.P. Barrett, Development of novel drugs for human African trypanosomiasis, Future Microbiol. 6 (2011) 677–691.

[8] P.G. Kennedy, Clinical features, diagnosis, and treatment of human African trypanosomiasis, Lancet Neurol. 12 (2) (2013) 186–194.

[9] World Health organization WHO, Mapping the Risk of Human African Trypanosomiasis, 2016. www.who.int/trypanosomiasis_african/country/risk_AFRo/en. (Accessed 27 February 2019).

[10] Trypanosomiasis, Human African (Sleeping Sickness), 2016. www.who.int/mediacentre/factsheets/fs259/en/, (Accessed 27 February 2019).

[11] J. Rodgers, Human African trypanosomiasis, chemotherapy and CNS disease, J. Neuroimmunol. 211 (2009) 16–22.

[12] F. Olmo, C. Rotger, I. Ramirez-Macias, L. Martinez, C. Marin, L. Carreras, K. Urbanova, M. Vega, G. Chaves-Lemaur, A. Sampedro, M.J. Rosales, M. Sanchez-Moreno, A. Costa, Synthesis and biological evaluation of N,N'-
squaramides with high in vivo efficacy and low toxicity: toward a low-cost drug against Chagas disease, J. Med. Chem. 57 (3) (2014) 987—999.

[13] J.J. Nogueira Silva, W.R. Pavanelli, F.R. Salazar Gutierrez, F.C. Alves Lima, A.B. Ferreira da Silva, J. Santana Silva, D. Wagner Franco, Complexation of the anti-trypanosoma cruzi drug benznidazole improves solubility and efficacy, J. Med. Chem. 51 (14) (2008) 4104—4114.

[14] S. Garcia, C.O. Ramos, J.F.V. Senra, F. Vilas-Boas, M.M. Rodrigues, A.C. Campos-de-Carvalho, R. Ribeiro-dos- Santos, M.B.P. Soares, Treatment with benznidazole during the chronic phase of experimental Chagas’ disease decreases cardiac alterations, Antimicrob. Agents Chemother. 49 (4) (2005) 1521—1528.

[15] S. Russell, R. Rahmani, A.J. Jones, H.L. Newson, K. Neilde, I. Cotillo, M. Rahmani Khajouei, L. Ferrins, S. Qureishi, N. Nguyen, M.S. Martinez-Martinez, D.F. Weaver, M. Kaiser, J. Riley, J. Thomas, M. De Rycker, K.D. Read, G.R. Flematti, E. Ryan, S. Tanghe, A. Rodriguez, S.A. Charman, A. Kessler, V.M. Avery, J.B. Baell, M.J. Piggott, Hit-to-lead optimization of a novel class of potent, broad-spectrum trypanosomacides, J. Med. Chem. 59 (21) (2016) 9686—9720.

[16] J.A. Urbina, R. Docampo, Specific chemotherapy of Chagas disease: controversies and advances, Trends Parasitol. 19 (11) (2003) 495—501.

[17] J.A. Castro, M.M. de Mecca, L.C. Bartel, Toxic side effects of drugs used to treat Chagas’ disease (American trypanosomiasis), Hum. Exp. Toxicol. 25 (8) (2006) 471—479.

[18] H.B. Tatipaka, J.R. Gillespie, A.K. Chatterjee, N.R. Norcross, M.A. Hulverson, R.M. Ranade, P. Nagendar, S.A. Creason, J. McQueen, N.A. Duster, A. Nagle, F. Supek, V. Molteni, T. Wenzler, R. Brun, R. Glynne, F.S. Buckner, M.H. Gelb, Substituted 2-henylimidazopyridines: a new class of drug leads for human African trypanosomiasis, J. Med. Chem. 57 (2014) 828—835.

[19] S. Chtita, M. Ghamali, M. Larif, R. Hmamouchi, M. Bouachrine, T. Lakhlifi, Quantitative structure—activity relationship studies of anticancer activity for Isatin (1H-indole-2,3-dione) derivatives based on density functional theory with electronic and topological descriptors, Int. J. Quant. Struct. Prop. Relatsh. 2 (2) (2017) 90—115.

[20] S. Chtita, M. Ghamali, R. Hmamouchi, B. Elidrissi, M. Bourass, M. Larif, M. Bouachrine, T. Lakhlifi, Investigation of antileishmanial activities of acridines derivatives against promastigotes and amastigotes form of parasites.
using quantitative structure activity relationship analysis, Adv. Phys. Chem. (2016) 1–16.

[21] Adamo, Baron, Parac & Grimme Gaussian 03 2003, 2000.

[22] ChemBioOffice, PerkinElmer Informatics, 2010. www.cambridgesoft.com. (Accessed 27 February 2019).

[23] XLSTAT, Software, XLSTAT Company, 2013. www.xlstat.com. (Accessed 27 February 2019).

[24] J.C. Dearden, M.T.D. Cronin, K.L.E. Kaiser, How not to develop a QSAR/QSPR, SAR QSAR Environ. Res. 20 (3-4) (2009) 241–266.

[25] J. Henseler, C. Ringle, R. Sinkovics, The use of partial least squares path modeling in international marketing, Adv. Int. Market. 20 (2009) 277–320.

[26] J.F. Hair, C.M. Ringle, Sarstedt, Partial least squares structural equation modeling: rigorous applications, better results and higher acceptance, Long. Range Plan. 46 (2013) 1–12.

[27] J.F. Hair, C.M. Ringle, M. Sarstedt, PLS-SEM: indeed a silver bullet, J. Mark. Theory Pract. 19 (2) (2011) 139–152.

[28] W.G. Zikmund, Business Research Methods, sixth ed., Harcourt College Publishers, Fort Worth, 2000.

[29] D.S. Moore, W.I. Notz, M.A. Flinger, The Basic Practice of Statistics, sixth ed., W. H. Freeman and Company, New York, NY, 2013.

[30] E.L. Lehmann, G. Casella, Theory of Point Estimation, second ed., Springer, New York, 1998.

[31] R.M. O’Brien, A caution regarding rules of thumb for variance inflation factors, Qual. Quantity 41 (2007) 673–690.

[32] K. Roy, On some aspects of validation of predictive quantitative structure–activity relationship models, Expert Opin. Drug Discov. 2 (12) (2007).

[33] D.W. Osten, Selection of optimal regression models via cross validation, J. Chemom. 2 (1998) 39–48.

[34] S. Ekins, G. Bravi, S. Binkley, J.S. Gillespie, B.J. Ring, J.H. Wikel, S.A. Wrighton, Three and four-dimensional-quantitative structure activity relationship (3D/4D-QSAR) analyses of CYP2C9 inhibitors, Drug Metab. Dispos. 28 (2000) 994–1002.
[35] OECD Guidance Document on the Validation of QSAR Models, Organization for Economic Co-operation & Development, Paris, 2007.

[36] L. Eriksson, J. Jaworska, A.P. Worth, M.T.D. Cronin, R.M. McDowell, P. Gramatica, Methods for reliability & uncertainty assessment & for applicability evaluations of classification & regression-based QSARs, Environ. Health Persp. 111 (10) (2003) 1361–1375.

[37] P. Gramatica, Principles of QSAR models validation: internal & external, QSAR Comb. Sci. 26 (5) (2007) 694–701.

[38] G.E. Batista, D.F. Silva, How K-Nearest Neighbor Parameters Affect its Performance, Argentine Symposium on Artificial Intelligence, Instituto de Ciencias Matemáticas de Computación, Sao Carlos – SP – Brasil, 2009, pp. 1–12.

[39] T.I. Netzeva, A.P. Worth, T. Aldenberg, R. Benigni, M.T.D. Cronin, P. Gramatica, J.S. Jaworska, S. Kahn, G. Klopman, C.A. Marchant, G. Myatt, N. Nikolova-Jeliazkova, G.Y. Patlewicz, R. Perkins, D.W. Roberts, T.W. Schultz, D.T. Stanton, J.J.M. Van De Sandt, W. Tong, G. Veith, C. Yang, Current status of methods for defining the applicability domain of (quantitative) structure–activity relationships, Altern. Lab. Anim. 33 (2) (2005) 155–173.

[40] J.C. Dearden, The history & development of QSARs, Int. J. Quantit. Struct. Prop. Relatsh. 1 (1) (2016).

[41] S. Zhang, A. Golbraikh, S. Oloff, H. Kohn, A. Tropsha, A novel Automated lazy learning QSAR (ALL-QSAR) Approach: method development, Applications, and virtual screening of chemical databases using validated ALL-QSAR models, J. Chem. Inf. Model. 46 (5) (2006) 1984–1995.

[42] K. Roy, S. Kar, R. Narayan Das, A Primer on QSAR/QSPR Modeling: Fundamental Concepts, Theoretical and Computational Chemistry, 2015.

[43] J. Holder, Ye. Lin, E.J. David, C. Cecil, C.C. Chappelow, A quantum-mechanical QSAR model to predict the refractive index of polymer matrices, QSAR Comb. Sci. 25 (10) (2006) 905–911.

[44] A.K. Srivastava, N. Shukla, A. Pandey, A. Srivastava, QSAR based modeling on a series of α-hydroxy amides as a novel class of bradykinin B1 selective antagonists, Journal of Saudi Chemical Society 15 (3) (2011) 215–220.

[45] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug
discovery and development settings, Adv. Drug Deliv. Rev. 6 (1–3) (1997) 3–25.

[46] S.C. Moldoveanu, V. David, Chapter 13 - Solvents, Buffers, and Additives Used in the Mobile Phase, Selection of the HPLC Method in Chemical Analysis, 2017, pp. 393–450.

[47] J. Isac-García, J.A. Dobado, F.G. Calvo-Flores, H. Martínez-García, Chapter 5 - Determining Physical and Spectroscopic Properties, Experimental Organic Chemistry, Academic Press, 2016, pp. 145–175.

[48] C. Hansch, C. Selassie, Quantitative structure—activity relationship — a historical perspective and the future, in: Reference Module in Chemistry, Molecular Sciences and Chemical Engineering, Comprehensive Medicinal Chemistry II, 4, 2007, pp. 43–63.

[49] A.A. Oliferenko, P.V. Oliferenko, J.G. Huddleston, R.D. Rogers, V.A. Palyulin, N.S. Zefirov, A.R. Katritzky, Theoretical scales of hydrogen bond Acidity and basicity for Application in QSAR/QSPR studies and drug design, J. Chem. Inf. Comput. Sci. 44 (3) (2004) 1042–1055.

[50] K. Fukui, Role of frontier orbitals in chemical reactions, Science 218 (1982) 747–754.

[51] S. Chtita, M. Larif, M. Ghamali, M. Bouachrine, T. Lakhlifi, 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, J. Taibah Univ. Sci. 9 (2) (2015) 143–154.

[52] V.H. Masand, N.N.E. El-Sayed, D.T. Mahajan, A.G. Mercader, A.M. Alafeefy, I.G. Shibi, QSAR modeling for anti-human African trypanosomiasis activity of substituted 2-Phenylimidazopyridines, J. Mol. Struct. 1130 (2017) 711–718.

[53] K. Roy, R.N. Das, P. Ambure, R.B. Aher, Be aware of error measures. Further studies on validation of predictive QSAR models, Chemometr. Intell. Lab. Syst. 152 (2016) 18–33.