

## A Quantitative Structure-Activity Relationships (QSAR) Study of Piperine Based Derivatives with Leishmanicidal Activity

Edilson Beserra de Alencar Filho<sup>a,\*</sup>, and Joseph Wallace Castro Silva<sup>a</sup>

<sup>a</sup>College of Pharmacy, Federal University of San Francisco Valley (UNIVASF), Av. José de Sá Maniçoba, Petrolina, Pernambuco, 56304-205, Brazil.

Article history: Received: 08 August 2016; revised: 15 December 2016; accepted: 18 January 2017. Available online: 30 March 2017. DOI: <http://dx.doi.org/10.17807/orbital.v9i1.893>

**Abstract:** Leishmaniasis is a parasitic disease which represents a serious public health problem in developing countries. It is considered a neglected tropical disease, for which there is little initiative in the search for therapeutic alternatives by pharmaceutical industry. Natural products remain a great source of inspiration for obtaining bioactive molecules. In 2010, Singh and co-workers published the synthesis and *in vitro* biological activity of piperoyl-aminoacid conjugates, as well as of piperine, against cellular cultures of *Leishmania donovani*. The piperine is an alkaloid isolated from *Piper nigrum* that has many activities described in the literature. In this work, we present a Quantitative Structure-Activity Study of piperine derivatives tested by Singh and co-workers, aiming to highlight important molecular features for leishmanicidal activity, obtaining a mathematical model to predict the activity of new analogs. Compounds were submitted to a geometry optimization computational procedure at semiempirical level of quantum theory. Molecular descriptors for the set of compounds were calculated by E-Dragon online platform, followed by a variable selection procedure using Ordered Predictors Selection algorithm. Validation parameters obtained showed that a good QSAR model, based on multiple linear regression, was obtained ( $R^2 = 0.85$ ;  $Q^2 = 0.69$ ), and the following conclusions regarding the structure-activity relationship were elucidated: Compounds with electronegative atoms on different substituent groups of analogs, absence of unsaturation on lateral chain, presence of ester instead of carboxyl, and large volumes (due the presence of additional aromatic rings) trends to increase the activity against promastigote forms of leishmania.

**Keywords:** E-Dragon; leishmanicidal activity; ordered predictors selection; piperine analogs; QSAR study

### 1. INTRODUCTION

Leishmaniasis is a disease caused by infection of the *Leishmania* genus parasites, which are found in parts of the tropics, subtropics, and southern Europe [1]. The infection is spread by the bite of phlebotomine sand flies and there are known three different types of clinical manifestations, which are: cutaneous, mucocutaneous and visceral [2]. Despite an estimate of mortality of 20,000–40,000 deaths per year for the visceral form [2], it is considered a neglected tropical disease, for which there is little initiative in the search for therapeutic alternatives by pharmaceutical industry. Actually, the treatment of leishmaniasis is performed with some drugs such as liposomal amphotericin B (for visceral leishmaniasis) and pentavalent antimonial compound sodium stibogluconate (Pentostam) (for cutaneous leishmaniasis), although there are still no preventive medications or vaccinations for the disease

[1]. These drugs are considered toxic, expensive and frequently ineffective [3], justifying the search for new molecules with leishmanicidal potential.

In the drug design process, biodiversity continues to be an important source of bioactive compounds, which provides molecules that have a direct action, as various anticancer and antimicrobial compounds [4], or serve as inspiration for the creation of new chemical entities. Piperine is an alkaloid found in *Piper nigrum* grains, a famous species reported from the tropical and subtropical regions of India [5]. Several biological activities are attributed to this molecule as antimicrobial [6], anti-inflammatory [7] and antiparasitic [8]. Based on piperine scaffold, in 2010 Singh and co-workers [9] published the synthesis and *in vitro* biological activity of piperoyl-aminoacid conjugates, as well as of piperine, against cellular cultures of *L. donovani*. In this work it was reported that the synthetic compounds were more active than

\*Corresponding author. E-mail: [edilson.beserra@univasf.edu.br](mailto:edilson.beserra@univasf.edu.br)

pure piperine ( $IC_{50} = 0.473$  mM for most active analogue and  $IC_{50} = 2.558$  mM for piperine), with values comparable to a known drug, mitefosine ( $IC_{50} = 0.033$  mM) [9]. That species cause the visceral leishmaniasis, which is the most severe form also known popularly as “kala-azar” [9].

Quantitative Structure-Activity Relationships study (QSAR) is a computationally viable approach that assumes the existence of relationships between the activity of a series of compounds (e.g. biological, toxic or environmental activities) and molecular descriptors, which are numerical values that feature different molecular properties [10]. These descriptors can be classified in general as: 0D descriptors, as the constitutional (molecular weight) and count descriptors (number of a type of atom); 1D descriptors, as fragment counts (number of a specific part of molecular structure); 2D descriptors, which use the atoms and connection information; 3D descriptors, based on static molecular conformations; 4D descriptors, which take in account conformational sampling profile; 5D (based on ligand-receptor interaction) and finally 6D descriptors (with an solvation scheme) [11]. Basically, the different QSAR study nomenclatures depend on the nature of the descriptor used. However, is important to note that there is not a better strategy than another, since the generated model is robust, predictive and help in the design of new potentially more active compounds. Thus, the final goal of QSAR analysis applied to pharmaceutical and biological sciences is to provide an equation relating the selected molecular descriptors with specific activity, allowing the design of compounds guided by the descriptors highlighted as well as predictions about trends of activity, saving cost and time in the drug development process [12].

Descriptors can be selected by *a priori* knowledge of the molecular characteristics involved on the mechanism of action, as well as through an initial calculation of hundreds or thousands of descriptors, followed by an appropriated variable selection algorithm (systematic search, genetic algorithm, incremental search, etc.). Chemometrical approaches are very useful in the QSAR modeling, providing statistical methods to construction of the multivariate models [13]. In general it can be made by Multiple Linear Regression (MLR) method, which conserves original nature of descriptors but is limited by the amount descriptors (ratio between compounds/descriptors must be  $> 5$ ) [14]. On the other hand, other statistical tools to QSAR modeling are based on Principal Component Analysis (PCA),

technique in which a large number of data, represented in Cartesian coordinates in a multidimensional space, is reduced to a number of principal components that store much of the information contained in the original axes [13]. Principal Component Regression (PCR) is a calibration scheme that is constructed similarly to MLR, using principal components as variables. Partial Least Squares (PLS) is another PCA based approach that considers the best relationships between biological activity and new axes in the modeling [14][15].

Considering biological data relative to *in vitro* leishmanicidal activity, provided by Singh and co-workers, in this paper we presented a 2D-QSAR study involving the seventeen piperoyl-aminoacid analogues. The goal of this work is to highlight molecular features important to leishmanicidal activity, obtaining a mathematical model to predict the activity new analogs, saving cost and time in the lead search and optimization.

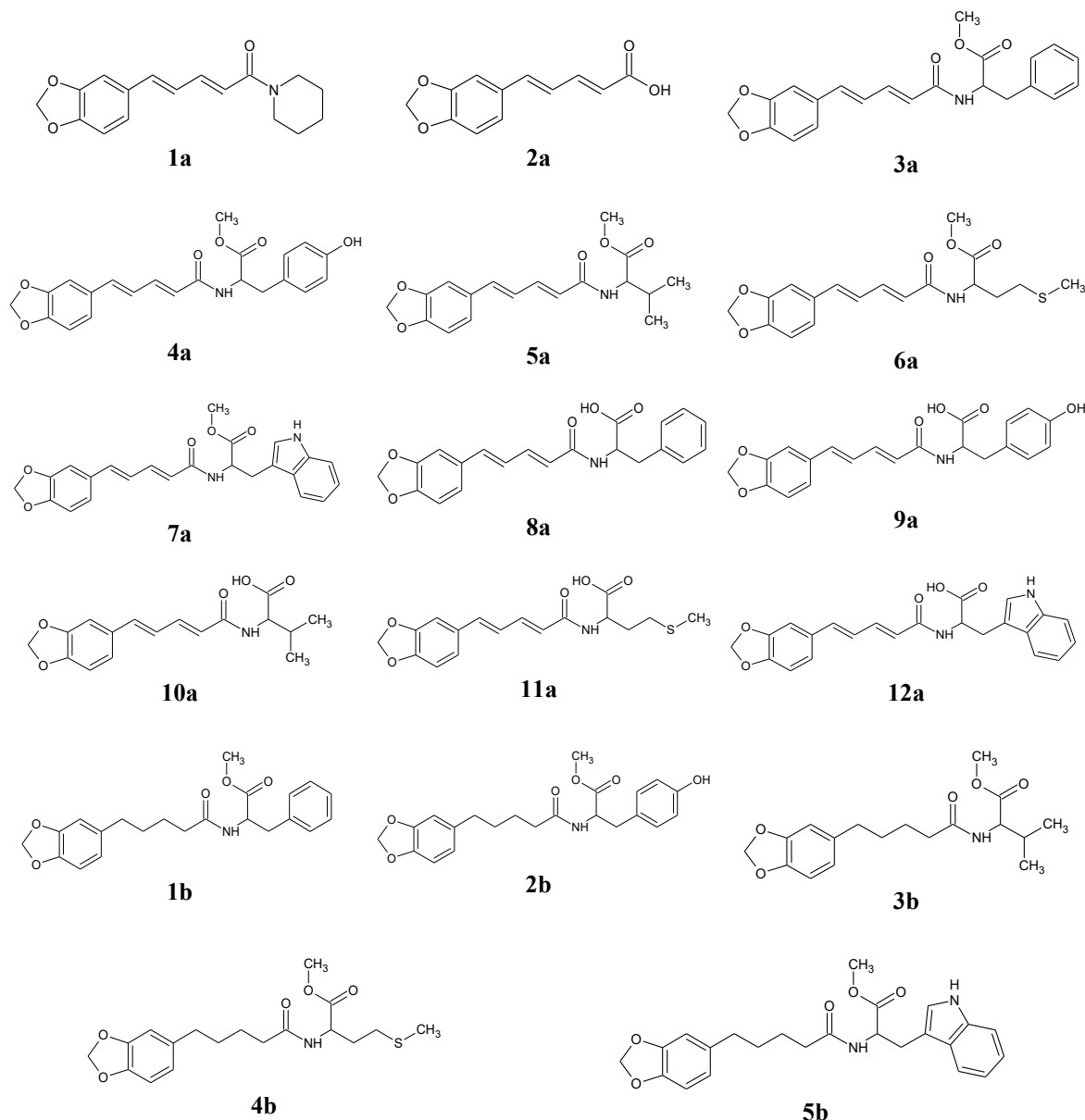
## 2. MATERIAL AND METHODS

The chemical structures of seventeen piperoyl-aminoacids conjugates (Figure 1) related by Singh and co-workers (2012) were constructed in ACD/ChemSketch program ([www.acdlabs.com](http://www.acdlabs.com)), saved in .mol format, and submitted to a geometry optimization, using the semi-empirical PM3 level of calculation [16] increasing the number of optimization steps until complete optimization, at Arguslab 4.0.1 software [17][18], in a Pentium Core i7 computer. The optimized geometries were upload on E-Dragon online platform to the calculation of 1.666 molecular descriptors of diverse nature, divided on 20 logical blocks (topological, constitutional, 2D autocorrelations, etc.) [19]. Biological activities on promastigote forms of *Leishmania* were converted from  $IC_{50}$  to  $-\text{Log}(IC_{50})$  ( $pIC_{50}$ ), reducing standard deviation and providing the highest numerical values for the most active compounds.

The matrix of descriptors was submitted to a variable selection procedure using Ordered Predictors Selector (OPS) algorithm [20] in QSAR modeling software [21]. This method is based on initial informative vector that reorder the original matrix, putting in the first columns the variables that have the greatest relationship with the activities presented (e.g. correlogram), followed by a systematic investigation of PLS regression models with the aim of finding the most relevant set of variables by comparing the cross-validation parameters of the models obtained [20]. No

training and test sets were defined due the relatively small number of compounds, thus allowing obtaining a greater capability for the QSAR model [22]. In this way, this study represents an initial direction for optimization leishmanicidal analogues of piperine. Considering the number of variables selected, MLR procedures were conducted using Build QSAR

program [23] providing good validation parameters, assessed by correlation matrix between variables, SPRESS value, coefficient of determination of calibration ( $R^2$ ), coefficient of determination of leave-one-out cross validation ( $Q^2$ ), correlation coefficient ( $r$ ), F-test value and standard deviation of predicted values ( $s$ ) [24].



**Figure 1.** 2D structures of seventeen piperoyl-aminoacid derivatives.

### 3. RESULTS AND DISCUSSION

During the optimization process, the few degrees of freedom allowed a relatively fast and secure conformational modeling, especially in molecules 1a-12a. No imaginary frequencies were observed in

optimized structures, which ensure obtaining of a minimum geometry.

Molecular descriptors selected after OPS algorithm and pIC50 values can be visualized on Table 1. Correlation matrix (Table 2) between selected

variables shows that they are not correlated ( $r^2 < 0.6$ ). MLR modeling provided the equation below (1), whose validation parameters can be visualized on Table 3.

$$pIC_{50} = -2.0524(MATS4v) - 5.9201(HATS3p) + 0.3802(Mor11p) + 3.1844 \quad (1)$$

**Table 1.**  $pIC_{50}$  values and molecular descriptors selected after OPS algorithm.

| Molecule | $pIC_{50}$ | MATS4v | HATS3p | Mor11p |
|----------|------------|--------|--------|--------|
| 1a       | 2.592      | -0.187 | 0.156  | -0.128 |
| 2a       | 2.755      | -0.143 | 0.108  | 0.06   |
| 3a       | 3.078      | -0.235 | 0.12   | 0.245  |
| 4a       | 3.101      | -0.229 | 0.107  | 0.079  |
| 5a       | 3.057      | -0.307 | 0.113  | -0.11  |
| 6a       | 3.084      | -0.253 | 0.099  | -0.277 |
| 7a       | 3.068      | -0.209 | 0.116  | 0.337  |
| 8a       | 3.032      | -0.233 | 0.133  | 0.393  |
| 9a       | 3.118      | -0.228 | 0.127  | 0.374  |
| 10a      | 2.987      | -0.312 | 0.135  | -0.192 |
| 11a      | 2.882      | -0.253 | 0.114  | -0.213 |
| 12a      | 2.989      | -0.204 | 0.124  | 0.419  |
| 1b       | 3.174      | -0.235 | 0.108  | 0.719  |
| 2b       | 3.270      | -0.229 | 0.101  | 0.487  |
| 3b       | 3.141      | -0.307 | 0.105  | 0.124  |
| 4b       | 3.216      | -0.253 | 0.096  | 0.059  |
| 5b       | 3.325      | -0.209 | 0.106  | 0.573  |

**Table 2.** Correlation matrix between selected descriptors.

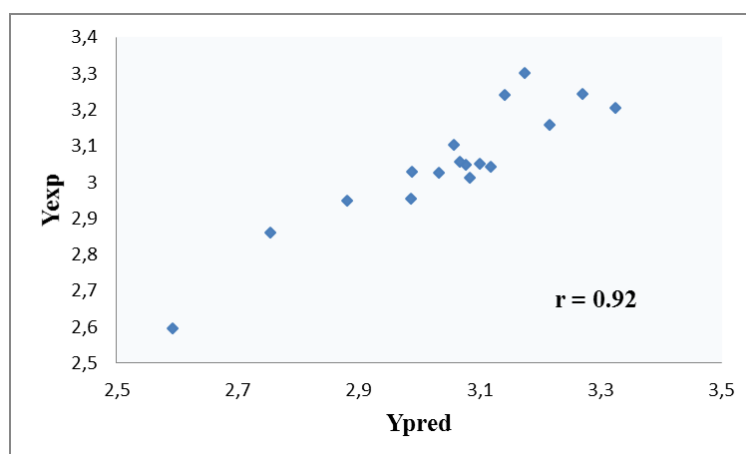
| Descriptor | MATS4v | HATS3p | Mor11p |
|------------|--------|--------|--------|
| MATS4v     | 1      | 0.128  | 0.334  |
| HATS3p     | 0.128  | 1      | 0.136  |
| Mor11p     | 0.334  | 0.136  | 1      |

**Table 3.** Validation parameters of QSAR model

| s    | SPRESS | $R^2$ | $Q^2$ | F     |
|------|--------|-------|-------|-------|
| 0.16 | 0.11   | 0.85  | 0.69  | 23.81 |

Validation parameters show that a good QSAR model was obtained. Standard deviation of predicted values is less than for experimental (0.18). Standard deviation of Prediction Error Sum of Squares (SPRESS) presented value close to zero. F-test also represents successful, considering the reference value  $F(3,13) = 3.34$  with 95% of confidence, showing statistic relevance of  $R^2$  [24]. Values of  $R^2$  and  $Q^2$  were

higher than the minimum established in the literature (0.6 and 0.5 respectively) [25]. Correlation coefficient also demonstrate satisfactory value, indicating good linearity between experimental and predicted leishmanicidal activities ( $r = 0.92$ ). The graphic of correlation between experimental and predicted values, as well the calculated residues can be visualized on Figure 3 and Table 4.



**Figure 3.** Correlation between experimental ( $Y_{exp}$ ) and predicted ( $Y_{pred}$ ) values of  $pIC_{50}$ .

**Table 4.** Residues between experimental and predicted values of pIC<sub>50</sub>.

| Molecule | Y <sub>exp</sub> | Y <sub>pred</sub> | Y <sub>res</sub> |
|----------|------------------|-------------------|------------------|
| 1a       | 2.592            | 2.596             | -0.004           |
| 2a       | 2.755            | 2.861             | -0.106           |
| 3a       | 3.078            | 3.049             | 0.029            |
| 4a       | 3.101            | 3.051             | 0.050            |
| 5a       | 3.057            | 3.104             | -0.047           |
| 6a       | 3.084            | 3.012             | 0.072            |
| 7a       | 3.068            | 3.055             | 0.013            |
| 8a       | 3.032            | 3.025             | 0.007            |
| 9a       | 3.118            | 3.043             | 0.075            |
| 10a      | 2.987            | 2.953             | 0.034            |
| 11a      | 2.882            | 2.948             | -0.066           |
| 12a      | 2.989            | 3.028             | -0.039           |
| 1b       | 3.174            | 3.301             | -0.127           |
| 2b       | 3.270            | 3.242             | 0.028            |
| 3b       | 3.141            | 3.240             | -0.099           |
| 4b       | 3.216            | 3.158             | 0.058            |
| 5b       | 3.325            | 3.204             | 0.121            |

One of the great contributions of QSAR approach is the interpretation of the physical meaning of the descriptors used in the model, to establish qualitative relationships between chemical structure and activity. This is an important step in design and proposing potentially more active structures, beyond the predictability inherent in the model. Generally, the literature to compare individually the value of descriptor with the activity, but it is important to note that is the balance between the various descriptors of a molecule that will provide the final value of its activity. In this sense, the nature of selected descriptors and their relationships with leishmanicidal activity are presented below:

**MATS4v:** Moran autocorrelation of lag 4 weighted by Van der Waals volume. Expresses how the values atomic properties are correlated, separated by a topological distance (lag). The “v” indicates that the descriptor is weighted by Van der Waals volumes, at the topological distance of eight bonds [26]. The expression of this descriptor is:

$$I(d) = \frac{\frac{1}{A} \sum_{i=1}^A \sum_{j=1}^A \delta_{ij} \cdot (w_i - \bar{w}) \cdot (w_j - \bar{w})}{\frac{1}{A} \sum_{i=1}^A (w_i - \bar{w})^2} \quad (2)$$

where  $w_i$  is an atomic property (Van der Waals volume),  $\bar{w}$  is the average on the molecule,  $A$  is the number of atoms and  $d$  corresponds to the topological distance (lag) between the atoms.  $\delta_{ij}$  corresponds to

the Kronecker delta, which is 1 if  $d_{ij} = d$  and zero if not. The negative value for the regression coefficient on QSAR equation indicates that most negative values of this descriptor promote best activities. In this way, the presence of heteroatoms on carbon scaffold leads to negative correlations and trends to increase the activity (e.g. 5a, 6a, 3b, 4b).

**HATS3p:** Leverage-weighted autocorrelation of lag 3, weighted by atomic polarizabilities. Is a GETAWAY descriptor (GEometry, Topology, and Atom-Weights Assembly), that combines the tridimensional molecular geometry (provided by molecular influence matrix H) with the chemical information, using different schemes of atomic ponderation (mass, polarizability, electronegativity) [26]. In the case of HATS3p represents autocorrelation between atoms separated by a distance of three bonds, weighted by their polarizabilities. In QSAR equation, regression coefficient is negative, indicating that higher values correspond to lowest pIC<sub>50</sub>. In fact, higher values of this descriptor are in molecules with less account of electronegative “hard” atoms (N, O, S), less polarizable (e.g. 1a, 8a, 10a). On the other hand, compounds as 4a and 6a have more electronegative atoms, lowest values of descriptors and trends to higher activity for “a” series. Curiously, “b” series have relatively the lowest values of HATS3p. This can be explaining due the absence of conjugated double bonds, decreasing relative polarizability and providing best activities.

**Mor11p:** Represents a 3D-MoRSE descriptor (3D Molecule Representation of Structures based on Electron diffraction) [26]. These descriptors are

calculated using the same dispersion function used on electron diffraction technique. These descriptors are directly related to the form and volume occupied by a molecule, where higher values can represent the total occupation of the binding site on biological target. The equation shows that higher values for this descriptor increase the activity. Thus, can be observed that molecules with ring substituent, and thus larger volumes, have higher value of the descriptor and a tendency to increase the activity (7a, 9a, 1b, 2b, 5b). In opposition, compounds as 1a, 2a, 10a and 11a have relatively smaller volumes, which lowers the value of Mor1lp, presenting lower activity.

In connection with the considerations above, we can note that the compounds with ester group are most actives, especially those that have rings on the lateral substituent, which is in agreement with the discussions. The presence of bulky aromatic rings on active compounds can be related to an increase on lipophilicity and/or possibility of establishing pi-stacking interactions on biological target. On the other hand, the absence of conjugated bonds on lateral chain of the most active compounds can be related to a major conformational freedom, facilitating the molecular recognition on biological medium.

Compound **5b** confirms our discussions, being the most active of the series and bringing together all the characteristics presented: Absence of conjugated bonds in the side chain, presence of the ester group, presence of the aromatic ring (indole), which includes an electronegative atom (nitrogen).

These discussions on the structure-activity relationships, attributing physical meaning to the molecular descriptors, corroborate the robustness and predictability mathematically demonstrated by the validation parameters.

#### 4. CONCLUSION

Considering the results and discussions presented, we can establish that use of E-Dragon descriptors, followed by OPS variable selection method, provided a good final MLR model for predicting activity of piperine derivatives, showing molecular features to be exploited in getting more active molecules. Observing the structure of the compounds, the values of activities, the values and meaning of selected descriptors, we conclude that the compounds with ester group associated to an aromatic ring with electronegative atom increase the leishmanicidal activity. Additionally, absence of

unsaturations on lateral moiety of piperine analogues leads to higher potency, possibly due the conformational freedom on biological medium. The final mathematical model can be used to provide estimates of the biological activity for new designed analogs.

#### 5. ACKNOWLEDGMENTS

The authors thank to the CNPQ, CAPES, FACEPE, UNIVASF and Prof. Ph.D Mário L. A. A. Vasconcellos (LASOM/UFPB) for some computational facilities.

#### 6. REFERENCES AND NOTES

- [1] Clem, A. *J. Glob. Infect. Dis.* **2010**, 2, 124. [\[CrossRef\]](#)
- [2] Savoia, D. *J. Infect. Dev. Ctries.* **2015**, 9, 588. [\[CrossRef\]](#)
- [3] Croft, S. L.; Coombs, G. H. *Trends Parasitol.* **2003**, 19, 502. [\[CrossRef\]](#)
- [4] Kingston, D. G. I. *J. Nat. Prod.* **2011**, 74, 496. [\[CrossRef\]](#)
- [5] Khan, S.; Mirza, K. J.; Anwar, F.; Zainul Abidin M. *Environ. We Int. J. Sci. Tech.* **2010**, 5, 47. [\[Link\]](#)
- [6] Reddy, P. S.; Jamil, K.; Madhusudhan, P.; Anjani, G.; Das, B. *Pharm. Bio.* **2001**, 39, 236. [\[CrossRef\]](#)
- [7] Pradeep, C. R.; Kuttan, G. *Int. Immunopharmacol.* **2004**, 4, 1795. [\[CrossRef\]](#)
- [8] Ribeiro, T. S.; Freire-de-Lima, L.; Previato, J. O.; Mendonça-Previato, L.; Heise, N.; Lima, M. E. F. *Bioorg. Med. Chem. Letters.* **2004**, 14, 3555. [\[CrossRef\]](#)
- [9] Singh, I. P.; Jain, S. K.; Kau, A.; Singh, S.; Kumar, R.; Garg, P.; Sharma, S. S.; Arora, S. K. *Eur. J. Med. Chem.* **2010**, 45, 3439. [\[CrossRef\]](#)
- [10] Tavares, L. C. *Quim. Nova.* **2004**, 27, 631. [\[CrossRef\]](#)
- [11] Liu, M.; Wilairat, P.; Go, L.M. *J. Med. Chem.* **2001**, 44, 4443. [\[CrossRef\]](#)
- [12] Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz'min, V. E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terfloth, L.; Gasteiger, J.; Richard, A.; Tropsha, A. *J. Med. Chem.* **2014**, 57, 4977. [\[CrossRef\]](#)
- [13] Beebe, K. R.; Pell, R. J.; Seasholtz, M. B. *Chemometrics: A Practical Guide.* New York: Wiley, 1998.
- [14] Ferreira, M. *J. Braz. Chem. Soc.* **2002**, 13, 742. [\[CrossRef\]](#)
- [15] Wold, S.; Sjöström, M.; Eriksson, L. *Chemom. Intel. Lab. Sys.* **2001**, 58, 109. [\[CrossRef\]](#)
- [16] Stewart, J. J. P. *J. Comput. Chem.* **1989**, 10, 209. [\[CrossRef\]](#)
- [17] Thompson. M. A. ArgusLab 4.0. Planaria Software LLC: Seattle. WA. **2015**. [\[Link\]](#)
- [18] Thompson M. A. *ACS meeting.* **2004**, Philadelphia 172. CINF 42. PA.
- [19] Tetko, I. V.; Gasteiger, J.; Todeschini, R.; Mauri, A.; Livingstone, D.; Ertl, P.; Palyulin, V. A.; Radchenko, E. V.; Zefirov, N. S.; Makarenko, A. S.; Tanchuk, V. Y.; Prokopenko, V. V. *J. Comput. Aid. Mol. Des.* **2005**, 19, 453.

- [20] Teófilo, R. F.; Martins, J. P.; Ferreira, M. M. C. *J. Chemometr.* **2009**, *23*, 32. [\[CrossRef\]](#)
- [21] Martins, J. A. M.; Ferreira, M. M. C. *Quim. Nova* **2013**, *36*, 554. [\[CrossRef\]](#)
- [22] Silva, J. B. P.; Ramos, M. N.; Neto, B. B.; Melo, S.; Falcão, E. P. S.; Catanho, M. T. J. A. *J. Braz. Chem. Soc.* **2008**, *19*, 337. [\[CrossRef\]](#)
- [23] De Oliveira, D. B.; Gaudio, A. C. *Quant. Struct.-Act. Rel.* **2001**, *19*, 599. [\[CrossRef\]](#)
- [24] Gaudio, A. C.; Zandonade, E. *Quim. Nova* **2001**, *24*, 658. [\[CrossRef\]](#)
- [25] Kiralj, R.; Ferreira, M. M. C. *J. Braz. Chem. Soc.* **2009**, *20*, 770. [\[CrossRef\]](#)
- [26] Strand life science p. Ltda, Theory Descriptors. 2007. [\[CrossRef\]](#)