Pyrimidine Derivatives: QSAR Studies of Larvicidal Activity against *Aedes aegypti*

Zenaide S. Monte,Daniela M. A. F. Navarro,Júlio C. R. O. F. Aguiar,Daniela M. A. F. Navarro,Júlio C. R. O. F. Aguiar,
Jessica S. Nascimento,Marcus T. Scotti,Luciana Scotti,Renata P. C. Barros,Aline C. S. Santos,Valéria R. A. Pereira,Emerson P. S. Falcão and Sebastião J. Melo* 

*Programa de Pós-Graduação em Ciências Farmacêuticas e em Ciências Biológicas, Universidade Federal de Pernambuco, Avenida Professor Morais Rego 1235, 50740-560 Recife-PE, Brazil  
Departamento de Química Fundamental, Universidade Federal de Pernambuco, 50740-560 Recife-PE, Brazil  
Programa de Pós-Graduação em Produtos Naturais e Sintéticos Bioativos, Universidade Federal da Paraíba, 58051-900 João Pessoa-PB, Brazil  
Centro de Pesquisas Aggeu Magalhães, Fundação Oswaldo Cruz, 50670-420 Recife-PE, Brazil  
Núcleo de Nutrição do Centro Acadêmico de Vitória (CAV), Universidade Federal de Pernambuco, 55608-680 Vitória de Santo Antão-PE, Brazil

The present study investigated the activity of pyrimidine derivatives against *Aedes aegypti*. Two compounds, 3c and 3d, showed excellent larvicide activity. Additionally, quantitative structure-activity relationship (QSAR) models were built using multiple-linear regression and partial least squares with descriptors generated from Dragon and VolSurf+ software, respectively. The best model is obtained with multiple linear regression (MLR), leading to a robust model. Moreover, the QSAR model is validated by means of some internal validation techniques in order to check its reliability, quality and robustness for predicting the larvicidal activity against *A. aegypti*. The models confirmed that the three-dimensional structure of molecules, steric properties, hydrophobic polar surface area, log partition (logP) and a simple pattern of substituent groups as methyl, methoxy, and succinimide in the pyrimidine derivatives are responsible for the larvicidal activity of the pyrimidine derivatives. Even more, the activity decreases by an electron-withdrawing group in R and increases when it is replaced by an aromatic ring activator group. These findings will aid in further studies of new pyrimidine derivatives active against *Aedes aegypti*.

**Keywords:** pyrimidine derivative, *Aedes aegypti*, QSAR model, robust model, MLR model

**Introduction**

Mosquitoes are a means of transmission of many neglected diseases, with millions of people being threatened vector-borne in the world, particularly in tropical and subtropical regions. According to Gorle et al., these diseases affected chiefly the tropical and subtropical regions of countries that resist chemical vector control programs because the population refused to prevent mosquito control by using chemical treatment with synthetic insecticides. According to the World Health Organization (WHO), in recent years, the transmission of dengue hemorrhagic fever, zika virus, dengue fever and chikungunya has increased dramatically in those regions that are conducive to mosquito proliferation, as these mosquitoes adapt in environments characterized by irrigation systems and heavy rains. Chikungunya produces fever and rheumatic pain, interfering with people’s quality of life for days, months or even years in more serious cases. Zika presents headache, irritation of the skin, redness of the eyes or vomiting, fatigue, fever, chills, loss of appetite or sweating. In some cases, Zika may cause paralysis (known as Guillain-Barré syndrome) and, in pregnant women, there may be hemorrhagic dengue: damp, pale and cold skin, as well as a decrease in blood pressure, high fever and malaise up to three days after the mosquito bite. Unfortunately, the use of synthetic insecticides
to combat the disordered growth of *Aedes aegypti* has not been effective in combating these mosquitoes since they are becoming resistant to conventional poisons, as well as increasing environmental problems and presenting serious damage to human health.27-19 Twenty-seven pyrimidine derivatives have already been reported in the literature by our research group.20,21 Because pyrimidines may have insecticidal activity,22 we investigated its biological activity against *A. aegypti*, with respect to the dengue vector. Additionally, quantitative structure-activity relationship (QSAR) studies were performed to understand the main physicochemical features responsible for the larvicidal activity of the studied compounds.

**Experimental**

**Maintenance and creation of the *Aedes* colony**

The beginning of the cycle was the hatching of *Aedes*’ eggs in glass cups containing enough distilled water to cover them completely and a little cat food to stimulate the hatching more quickly. The development of the larval cycle continued by changing water in the basins and adding more food. The end of one cycle and the beginning of the next cycle was accomplished by putting the larvae and pulp in a glass beaker with a little distilled water and food (cat food) into cages. The following week, the cups were removed and a blood repast was added to the liquid. New egg collection was conducted with filter paper and dried at room temperature of 27 ± 1 °C and 75% humidity. The mosquitoes were kept alive with a piece of cotton soaked with 10% sucrose solution.

**Experimental procedure for larvicidal bioassays**

Preliminary solubility tests were initially conducted for each sample aliquot to be tested in co-solvents such as ethanol, tween 80 or acetone, in order to select the co-solvent which best solubilized the sample in water. After this test, a stock solution concentration equal to 100 parts per million (ppm) was prepared by dissolving 5.0 mg of the compound to be tested in a volume of 0.7 mL of the chosen co-solvent.23,24 The contents were then transferred from the beaker into a 50 mL volumetric flask and the volume topped with distilled water. Preliminary larvicidal tests were performed in concentrations of 10, 50 and 100 ppm in order to observe the concentration range where the compound was most active. The tests were carried out in triplicate which, in each replica, 20 larvae of *A. aegypti* were used and larvicidal activity was observed after 24 and 48 h from the beginning of the test. The larvae were considered dead when they did not respond to the stimulus or did not emerge on the surface of the solution. Negative controls (solution containing only co-solvent and distilled water) were carried out simultaneously with the tests. For the determination of the lethal concentration (LC₅₀) values for 50% of the larval population, the data was obtained using Probit software25 with the statistical program StatPlus Pro 6.2.5.0,26 at a 95% confidence level.

**Evaluation of cytotoxic activity**

Macrophage cell lines RAW 264.7 were cultured in a complete Dulbecco’s modified Eagle medium (DMEM, 100 mg mL⁻¹ streptomycin, 100 U mL⁻¹ penicillin and 10% fetal calf serum (FCS)), (Cultilab, Campinas, São Paulo, Brazil). Cells were maintained at 37 °C, in a 5% CO₂ atmosphere. Macrophages were seeded at 10⁶ cells per well in 96-well plates and incubated for 24 h (37 °C and 5% CO₂). Compounds were then added to six different concentrations (6.25 to 200 μg mL⁻¹). Wells containing only the culture medium and cells (without treatment) were used as the negative control. Cells were incubated for 48 h. Then a MTT solution (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazoliumbromide (Sigma, St. Louis, MO, USA) at 5 mg mL⁻¹ in phosphate buffered saline (PBS) was added. The plates were incubated again for 2 h. The remaining culture medium and the unredced MTT were removed and 100 μL of dimethyl sulfoxide (DMSO) was added for solubilization of formazan. The amount of formazan was determined by measuring the absorbance at 570 nm. The assay was performed in triplicate. The cytotoxicity concentration (CC₅₀) was calculated by regression analysis with GraphPad Prism Software 5.0 (San Diego, CA, USA).27

**Statistical analysis**

Statistical analysis was performed using non-parametric tests. A simple linear regression test was performed to obtain CC₅₀.

**Molecular descriptors**

Initially, the SMILES (simplified molecular line entry system) code of the molecules was obtained, generating a single file, .smi format, of all molecules. This file was introduced into the software Standardizer, ChemAxon,28 to canonize structures, add hydrogens, perform aromatic form conversions, clean the molecular graph in three dimensions. The process uses a divide-and-conquer approach. The structure is split into small fragments which
are organized into a tree using connectivity information. Conformers generated for the initial structure (represented by the root node in the tree) are optimized. The tree building process uses a proprietary extended version of the Dreiding force field. Finally, the compounds were saved in SDF format that was used as input in two software programs to generate the molecular descriptors. The first program was the VolSurf+ v. 1.0.7, which produced the file containing the structures in 3D, by calculating molecular descriptors through the intrusion of these molecules to the molecular interaction fields (MIFs), using 4 different probes: amide nitrogen (N1), carbonyl oxygen (O), hydrophobic probe (DRY) and water probe (H2O). It was also calculated some molecular descriptors not derived from the field of molecular interaction (non-MIF), generating a total of 128 descriptors; for example, quantifiable descriptors such as hydrophobic-lipophilic balance, capacity factors, molecular size, hydrophobic and hydrophilic regions, amphilic moments and moments of energy interaction. Due to ease of use, understanding and interpretation of these descriptors, they were chosen the present study.

The second program used was Dragon v. 7.0.10, to calculate a total of 5255 molecular descriptors through 29 descriptor blocks. After calculating these descriptors, a data treatment was performed in which descriptors of constant variables were excluded for each descriptor block, besides the exclusion of those descriptors that had a high degree of correlation (r < 0.95).

Model generation

The QSAR models were developed using the partial least squares (PLS) method with VolSurf+ software. The PLS is based on linear regression, making it possible to extract and rationalize multivariate information to explain the maximum correlation between the descriptors, matrix x and matrix y, to then calculate a new set of orthogonal variables that had not yet been correlated, the latent variables (LVs). Whenever the number of variables is greater than the number of samples and multicollinearity occurs among independent variables, this method becomes appropriate. The self-scale pre-processing was performed and applied to all independent variables, by subtracting the mean by the values of each variable and then dividing the resulting values by the standard deviation. A variable influence in projection (VIP) plot was used to select the number of original variables. The VIP parameter has the ability to condense the importance of the variables in the PLS model and to quantify this importance by the coefficient values and indicate the contribution of each descriptor to the model.

The model performance was estimated by the variance explained in the coefficient of determination (r²) and the coefficient of determination in the cross-validation by leave-one-out (Qcv). The model and number of ideal LVs, that is, orthogonal linear combinations of the original variables, were determined by the highest value of Qcv. MobyDigs software was also used to calculate and generate regression models using a genetic algorithm, in which the models for the activity were constructed and internally validated using leave-one-out Qcv.

Results and Discussion

Insecticidal activities

Insecticidal resistance is the main problem in the control of the mosquito population in the world, principally for A. aegypti. In view of this, compounds of pyrimidine derivatives with potential anti-larvicidal activity were investigated to detect mortality of larvae in the presence of these compounds. The 27 compounds 5a-f and 6g-i screened for larvicidal activity against four instar larvae of A. aegypti are listed in Table 1. Considerable variation was observed in the susceptibility of the larvae to the different pyrimidines, indicating those which proved to be the most promising agents, with impressive larvicidal activity. Thus, among all the synthesized compounds, 3l and 5a showed highest mortality results (56.7 ± 15.4, 65.7 ± 15.9, respectively) and reached standard value at 100 ppm concentration at 72 h; the other 10 compounds 3a-d, 3m, 5a and 5g-d achieved ≤ 50% mortality level of larval mortality in concentration of 100 ppm. Among all the synthesized compounds, 3l and 5a showed highest mortality results (56.7 ± 15.4, 65.7 ± 15.9, respectively) and reached standard value at 100 ppm concentration at 72 h of exposure; other results were higher than the standard value at 100 ppm concentration. Impressive mortality was observed for 3c and 3d, with values 5.4 ± 1.8 and 4.4 ± 0.5, respectively. The mortality rates were considered at 100 ppm concentration at 72 h.

In parallel to the evaluation of larvicidal activity, the toxicity of the compounds on mammalian cells was determined, using the murine macrophage RAW 264.7. A variation in the activity of the pyrimidines was observed. Nine, of the 27 pyrimidines in Table 1 showed lower cytotoxicity, with high CC50 values, above 100 μM: 3a, 3b, 3f, 3j, 3k, 3m, 3n and 5a. Another ten, showed higher toxicity, with CC50 values below 50 μM: 3d, 3e, 3g, 3l, 3n, 3p, 3q, 5c and 5i. Intersecting the results obtained in the evaluation of larvicidal activity and mammalian cell toxicity, the pyrimidines which showed the best results, i.e., low toxicity in macrophages and high mortality rate at the
concentration of 100 ppm for *A. aegypti* larvae were: 3f, 3j, 3k and 5a.

**QSAR models**

In order to minimize errors in the studies by QSAR, Kubinyi recommends that the biological activity values should be standardized (inhibition concentration-IC$_{50}$, LC$_{50}$, lethal dose-LD$_{50}$, etc) and that the variation in the values of activity between the most active and least active compounds is at least one log unit. It can be observed that the difference of logarithmic units between the highest activity compound, 3d $p$LC$_{50} = 4.9$, and the lowest activity compound, 5a $p$LC$_{50} = 3.75$, is 1.15, in accordance with the basic principle of the QSAR study.

Regression analysis of the training set generated

<table>
<thead>
<tr>
<th>Compound</th>
<th>R$_1$</th>
<th>R$_2$</th>
<th>R$_3$</th>
<th>Concentration range / ppm</th>
<th>LC$_{50}$ / µM</th>
<th>Confidence interval / %</th>
<th>Standard error for LC$_{50}$</th>
<th>Cytotoxicity CC$_{50}$ / µM</th>
</tr>
</thead>
<tbody>
<tr>
<td>3a</td>
<td>H</td>
<td>NO$_2$</td>
<td>H</td>
<td>50-20</td>
<td>46.3</td>
<td>42.8-52.5</td>
<td>2.4</td>
<td>112.1</td>
</tr>
<tr>
<td>3b</td>
<td>H</td>
<td>H</td>
<td>NO$_2$</td>
<td>50-15</td>
<td>18.6</td>
<td>14.3-21.6</td>
<td>1.95</td>
<td>124.55</td>
</tr>
<tr>
<td>3c</td>
<td>OMe</td>
<td>NO$_2$</td>
<td>H</td>
<td>25-4</td>
<td>5.4</td>
<td>4.3-6.4</td>
<td>0.5</td>
<td>88.56</td>
</tr>
<tr>
<td>3d</td>
<td>OMe</td>
<td>H</td>
<td>NO$_2$</td>
<td>45-5</td>
<td>4.4</td>
<td>1.4-7.0</td>
<td>1.8</td>
<td>16.37</td>
</tr>
<tr>
<td>3e</td>
<td>C$<em>{4}$H$</em>{4}$N</td>
<td>NO$_2$</td>
<td>H</td>
<td>–</td>
<td>&gt; 100</td>
<td>–</td>
<td>–</td>
<td>33.29</td>
</tr>
<tr>
<td>3f</td>
<td>C$<em>{4}$H$</em>{4}$N</td>
<td>H</td>
<td>NO$_2$</td>
<td>–</td>
<td>&gt; 100</td>
<td>–</td>
<td>–</td>
<td>203.56</td>
</tr>
<tr>
<td>3g</td>
<td>C$<em>{4}$H$</em>{4}$N</td>
<td>H</td>
<td>Cl</td>
<td>–</td>
<td>&gt; 100</td>
<td>–</td>
<td>–</td>
<td>31.07</td>
</tr>
<tr>
<td>3h</td>
<td>OMe</td>
<td>H</td>
<td>Cl</td>
<td>–</td>
<td>&gt; 100</td>
<td>–</td>
<td>–</td>
<td>57.06</td>
</tr>
<tr>
<td>3i</td>
<td>NH$_2$</td>
<td>H</td>
<td>OMe</td>
<td>–</td>
<td>&gt; 100</td>
<td>–</td>
<td>–</td>
<td>82.69</td>
</tr>
<tr>
<td>3j</td>
<td>NH$_2$</td>
<td>NO$_2$</td>
<td>H</td>
<td>–</td>
<td>&gt; 100</td>
<td>–</td>
<td>–</td>
<td>188.06</td>
</tr>
<tr>
<td>3k</td>
<td>NH$_2$</td>
<td>H</td>
<td>NO$_2$</td>
<td>–</td>
<td>&gt; 100</td>
<td>–</td>
<td>–</td>
<td>264.78</td>
</tr>
<tr>
<td>3l</td>
<td>NO$_2$</td>
<td>NO$_2$</td>
<td>H</td>
<td>100-50</td>
<td>56.7</td>
<td>24.0-68.7</td>
<td>15.4</td>
<td>49.81</td>
</tr>
<tr>
<td>3m</td>
<td>NO$_2$</td>
<td>H</td>
<td>NO$_2$</td>
<td>50-20</td>
<td>31.4</td>
<td>28.3-34.9</td>
<td>1.7</td>
<td>131.47</td>
</tr>
<tr>
<td>3n</td>
<td>NO$_2$</td>
<td>H</td>
<td>OMe</td>
<td>–</td>
<td>&gt; 100</td>
<td>–</td>
<td>–</td>
<td>46.89</td>
</tr>
<tr>
<td>3o</td>
<td>Me</td>
<td>NO$_2$</td>
<td>H</td>
<td>40-5</td>
<td>17.0</td>
<td>14.1-19.4</td>
<td>1.4</td>
<td>175.85</td>
</tr>
<tr>
<td>3p</td>
<td>Me</td>
<td>H</td>
<td>NO$_2$</td>
<td>10-45</td>
<td>18.5</td>
<td>16.8-20.4</td>
<td>0.9</td>
<td>30.02</td>
</tr>
<tr>
<td>3q</td>
<td>Me</td>
<td>H</td>
<td>Cl</td>
<td>–</td>
<td>&gt; 100</td>
<td>–</td>
<td>–</td>
<td>17.39</td>
</tr>
<tr>
<td>3r</td>
<td>Cl</td>
<td>H</td>
<td>H</td>
<td>–</td>
<td>&gt; 100</td>
<td>–</td>
<td>–</td>
<td>87.8</td>
</tr>
<tr>
<td>5a</td>
<td>H</td>
<td>succinimide</td>
<td>H</td>
<td>120-70</td>
<td>65.7</td>
<td>41.0-105.3</td>
<td>15.9</td>
<td>136.1</td>
</tr>
<tr>
<td>5b</td>
<td>H</td>
<td>succinimide</td>
<td>H</td>
<td>70-20</td>
<td>48.5</td>
<td>36.8-109.3</td>
<td>13.64</td>
<td>52.36</td>
</tr>
<tr>
<td>5c</td>
<td>OMe</td>
<td>succinimide</td>
<td>H</td>
<td>35-10</td>
<td>11</td>
<td>7.0-13.6</td>
<td>1.9</td>
<td>17.32</td>
</tr>
<tr>
<td>5d</td>
<td>OMe</td>
<td>H</td>
<td>succinimide</td>
<td>40-10</td>
<td>12.99</td>
<td>9.5-15.5</td>
<td>1.64</td>
<td>128.85</td>
</tr>
<tr>
<td>5e</td>
<td>C$<em>{4}$H$</em>{4}$N</td>
<td>succinimide</td>
<td>H</td>
<td>–</td>
<td>&gt; 100</td>
<td>–</td>
<td>–</td>
<td>99.3</td>
</tr>
<tr>
<td>5f</td>
<td>C$<em>{4}$H$</em>{4}$N</td>
<td>succinimide</td>
<td>H</td>
<td>–</td>
<td>&gt; 100</td>
<td>–</td>
<td>–</td>
<td>84.4</td>
</tr>
<tr>
<td>6g</td>
<td>pyrrol-2-one</td>
<td>H</td>
<td>C$<em>{2}$H$</em>{6}$N</td>
<td>–</td>
<td>&gt; 100</td>
<td>–</td>
<td>–</td>
<td>60.94</td>
</tr>
<tr>
<td>6h</td>
<td>pyrrol-2-one</td>
<td>H</td>
<td>OMe</td>
<td>–</td>
<td>&gt; 100</td>
<td>–</td>
<td>–</td>
<td>56.03</td>
</tr>
<tr>
<td>6i</td>
<td>OMe</td>
<td>H</td>
<td>maleimide</td>
<td>–</td>
<td>&gt; 100</td>
<td>–</td>
<td>–</td>
<td>37.2</td>
</tr>
</tbody>
</table>

LC$_{50}$: lethal concentration; CC$_{50}$: cytotoxicity concentration.
equation 1, with TDB05v (3D Autocorrelations) and Mor04u (3D MoRSE) descriptors, which are able to explain almost 94.4% of the variance in antilarval activity.  

\[
pLC_{50} = -0.317 \pm 0.070 \times TDB05v - 0.218 \pm 0.558 \times \text{Mor04u} + 4.24 \pm 0.067 \  \quad (1) 
\]

\((n = 12; r^2 = 0.944; s = 0.102; F = 75.94; Q^2_{cv} = 0.920; \text{SDEP} = 0.106)\)

where \(n\) is the number of samples, \(r^2\) is the coefficient of determination, \(s\) is the mean square error, \(F\) is the Fisher function, \(Q^2_{cv}\) is the cross-validated \(r^2\), and \(\text{SDEP}\) is standard deviation error of prediction.

An analysis of Table 2 and Figure 1 indicates that each value represents the adjustment in relation to a line of the points that had been used for the calibration of the model. Equation 1, described above, shows that the value of the internal prediction (leave-one-out) coefficient \(Q^2_{cv}\) is considerable \((0.920)\), indicating how robust the model was. The value of \(F\) \((75.94)\) was highly expressive, with 95% significance and with 2 and 9 degrees of freedom, where the minimum required value is 4.26. In the conception of obtaining information from 3D (three-dimensional) atomic coordinates through the transformation technique used in electron diffraction studies, one of them Mor04u, is based on the 3D-morse descriptor (3D-molecule representation of structures based on electron diffraction) selected in equation 1. The descriptor TDB05v is a 3D-topological distance-based descriptor, which is based on Moreau-Broto’s 2D (two-dimensional) autocorrelation that portrays how a certain property is spread throughout a topological molecular structure, also encoding information about the separation space between two atoms.  

The larvicidal activity is closely related to the three-dimensional

<table>
<thead>
<tr>
<th>Compound</th>
<th>pLC_{50}^{\log} / M</th>
<th>pLC_{50} Dragon / M</th>
<th>Error Dragon</th>
<th>pLC_{50} Vols / M</th>
<th>Error Vols</th>
<th>pLC_{50} consensus / M</th>
<th>Error Consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td>3a</td>
<td>3.84</td>
<td>3.87</td>
<td>0.03</td>
<td>4.01</td>
<td>0.17</td>
<td>3.94</td>
<td>0.10</td>
</tr>
<tr>
<td>3b</td>
<td>4.23</td>
<td>4.01</td>
<td>-0.22</td>
<td>3.99</td>
<td>-0.24</td>
<td>4.00</td>
<td>-0.23</td>
</tr>
<tr>
<td>3c</td>
<td>4.81</td>
<td>4.85</td>
<td>0.04</td>
<td>4.54</td>
<td>-0.27</td>
<td>4.70</td>
<td>-0.11</td>
</tr>
<tr>
<td>3d</td>
<td>4.90</td>
<td>4.97</td>
<td>0.07</td>
<td>5.02</td>
<td>0.12</td>
<td>5.00</td>
<td>0.09</td>
</tr>
<tr>
<td>3l</td>
<td>3.81</td>
<td>3.87</td>
<td>0.06</td>
<td>3.79</td>
<td>-0.02</td>
<td>3.83</td>
<td>0.02</td>
</tr>
<tr>
<td>3m</td>
<td>4.06</td>
<td>3.98</td>
<td>-0.08</td>
<td>4.19</td>
<td>0.13</td>
<td>4.09</td>
<td>0.03</td>
</tr>
<tr>
<td>3o</td>
<td>4.29</td>
<td>4.23</td>
<td>-0.06</td>
<td>4.14</td>
<td>-0.15</td>
<td>4.19</td>
<td>-0.11</td>
</tr>
<tr>
<td>3p</td>
<td>4.25</td>
<td>4.38</td>
<td>0.13</td>
<td>4.59</td>
<td>0.34</td>
<td>4.49</td>
<td>0.23</td>
</tr>
<tr>
<td>5a</td>
<td>3.75</td>
<td>3.92</td>
<td>0.17</td>
<td>3.96</td>
<td>0.21</td>
<td>3.94</td>
<td>0.19</td>
</tr>
<tr>
<td>5b</td>
<td>3.88</td>
<td>3.95</td>
<td>0.07</td>
<td>3.67</td>
<td>-0.21</td>
<td>3.81</td>
<td>-0.07</td>
</tr>
<tr>
<td>5c</td>
<td>4.56</td>
<td>4.45</td>
<td>-0.11</td>
<td>4.50</td>
<td>-0.06</td>
<td>4.48</td>
<td>-0.09</td>
</tr>
<tr>
<td>5d</td>
<td>4.49</td>
<td>4.48</td>
<td>-0.01</td>
<td>4.58</td>
<td>0.09</td>
<td>4.53</td>
<td>0.04</td>
</tr>
</tbody>
</table>

~logarithm of the lethal concentration values for 50% of the larval population in values of experimental activity \((pLC_{50})\).

![Figure 1](image_url)  
**Figure 1.** Graph of the MLR model generated from the Dragon descriptors of equation 1, representing the values of experimental activity \((pLC_{50})\) versus predicted (leave-one-out cross validation) activity values.
structure of molecules as well as the steric properties of specific points of the molecule.

The PLS model selected with five LVs demonstrated the following statistical parameters: $Q_{cv}^2 = 0.741$ (Figure 2 and Table 2), standard deviation of errors of prediction (SDEP) = 0.190, $R^2 = 0.934$ and standard deviation of errors of calculation (SDEC) = 0.096.

The graphs of the $t_1$-$t_2$ scores and loadings with two LVs of the PLS model can be observed in Figure 3, showing a separation between the most active compounds (black points) and less active (white points). The white points to the left of the graph indicate the lower experimental activity ($pLC_{50}$) chemical structures, while the black points to the right of the graph depict the chemical space of the molecules with the highest $pLC_{50}$ values.

In the loading plot, (Figure 3b), the descriptors VolSurf+ that contributed more in the PLS model can be seen. The HSA (hydrophobic surface area) and log partition between cyclohexane and water (logPc-hex) contribute positively to the $pLC_{50}$ values. These descriptors are related to the hydrophobic characteristics of these molecules. The HSA is computed through the hydrophobic region of the molecule and the logPc-hex. The physical-chemical characteristics described by the HSA descriptors and logPc-hex are important for the discovery of new drugs since they provide pharmacokinetic properties in the initial phase, revealing molecules with greater chance of being transported through the circulatory system to the target tissues. Since the $LC_{50}$ values were obtained from the test against the larvae mosquito, these properties are related to potential toxicity. On the other hand, the descriptors PSA (polar surface area), PSAR (ratio between the PSA and the surface) and PHSAR (polar hydrophobic surface area), the ratio between the polar surface area PSA and the HSA, contribute negatively to $pLC_{50}$ values, therefore the more polar compounds are less active. Figure 4 shows the lipophilic and hydrophilic regions of the higher activity compound, $3d$, and the compound with the lower activity, $3l$. It can be observed that compound $3d$ has a larger lipophilic area and smaller hydrophilic surface area when compared with molecule $3l$, thus justifying its better activity.

Given the results and analyzing the most active and least active compounds, we can observe the influence of the substituents on the radicals $R_1$, $R_2$, and $R_3$. The PLS model classified that molecules with higher lipophilicity ($logPn$-hex) and HSA have higher activities and those with smaller PSA, smaller radius between PSAR and the smallest radius between the polar surface area and the PHSAR has low biological activity. Thus, we can analyze the substituents $R_1$, $R_2$, and $R_3$ to understand the relationship between chemical structure and biological activity.

First looking at the radicals $R_2$ and $R_3$ when substituted by the nitro group (NO$_2$), we see that when located at $R_3$, position para, there is a slight improvement in activity, comparing, for example, molecules $3d$ and $3c$, where the only difference is precisely the position of the group nitro. This slight improvement is explained by the decrease in PSAR and PHSAR when nitro is in the $R_3$ target position, Table 3. Similarly, it occurs when the substituent of $R_2$ and $R_3$ is succinimide, molecules $5c$ and $5d$, where substitution in $R_2$, meta position, is favored.

Now looking at the radical $R_1$, which has been replaced by OCH$_3$, CH$_3$, H and NO$_2$, we see that when there is a methoxyl (OCH$_3$) the activity profile of the molecule is better than when replaced by a methyl, molecules $3c$ and $3o$, for example. When substituted by methyl group there is a considerable decrease in PSA, PSAR and PHSAR (Table 3), causing the activity to be decreased, $pLC_{50}$ from 4.81 for

Figure 2. Plot of experimental activity values ($pLC_{50}$) versus the predicted values by cross-validation leave-one-out of the selected partial least squares (PLS) model.
molecule 3c and from 4.29 for molecule 3d. Comparing now the methoxyl with the nitro group, molecules 3c and 3l, we have that the methoxyl group has a higher lipophilicity due to the presence of methyl and there is also an increase in hydrophobic surface area, explaining why the molecule 3c is more active than molecule 3l, with 3l being one of the least active (pLCso = 3.81). Similar behavior occurs when R1 is replaced by hydrogen compared to methoxyl substitution, molecules 5c and 5a, the increase in logPc-hex and HSA (Table 3) will be greater with methoxyl, so the molecule will have better activity.

An important fact that should also be considered in the analysis is that the vector sum of the individual bond dipole moments will change the polar and nonpolar profile.

Figure 3. Graphs of the partial least squares (PLS) model. The (a) scores and (b) loadings plot with the two latent variables (LV1 vs. LV2).
of the molecule, and that certain geometries make the resulting dipole moment null, resulting in nonpolarity of the molecule.\textsuperscript{43-49}

A consensus model was obtained from calculating the average values of the two models generated by the combination of the MLR and PLS models with Dragon and VolSurf+ descriptors respectively (Table 2). The predicted activity was estimated by taking an average of the predicted pLC\textsubscript{50} from both QSAR methods. This procedure usually provides better prediction accuracy than the majority of individual models since errant predictions are dampened by the predictions from the other methods. The performance of the MLR model is better for all samples, except for compounds 3l and 5c. The consensus model shows almost the same performance as the MLR model and better performance than the PLS model. Finally, the consensus shows the following parameters: Q\textsuperscript{2} = 0.879, standard deviation of errors of prediction (SDEP) = 0.130.

Although the descriptors related to the electronic effect of the substituent groups were not selected, we also observed that there is an electronic pattern relating the chemical structure of the compounds with the biological activity. This electronic influence is specifically perceived in the radical R\textsubscript{1}, where there are variations of the substituent groups. When R\textsubscript{1} is replaced by a ring deactivator group, activity decreases, as in the case of the nitro group which is a strong ring deactivator, as does molecule 3l which has one of the smallest biological activities. When R\textsubscript{1} is replaced by an aromatic ring activator group, electro donor, biological activity also increases.

We can compare, for example, molecules 3l, 3o, 3c and 3j to 3l as already mentioned have an aromatic ring deactivating group, whereas molecules 3o, 3c and 3j have a ring activating group. Molecule 3c has a methoxy in R\textsubscript{1} which is a moderate activator leading to a good biological activity of the molecule, 3d has a methyl which is a weak activator of the aromatic ring and there is a decrease in biological activity when compared to 3c, and molecule 3j has an amine at R\textsubscript{1} which is a strong ring activator, but it greatly elevates molecule’s hydrophilia and therefore leaves it inactive.

Table 3. Values of descriptors selected in partial least squares (PLS) for the most active and least active molecules

<table>
<thead>
<tr>
<th>Compound</th>
<th>logP\textsubscript{c-hex}\textsuperscript{a}</th>
<th>HSA\textsuperscript{b}</th>
<th>PSA\textsuperscript{c}</th>
<th>PSAR\textsuperscript{d}</th>
<th>PHSAR\textsuperscript{e}</th>
<th>pLC\textsubscript{50}\textsuperscript{f}</th>
</tr>
</thead>
<tbody>
<tr>
<td>3d</td>
<td>2.818</td>
<td>385.546</td>
<td>123.65</td>
<td>0.242</td>
<td>0.320</td>
<td>4.90</td>
</tr>
<tr>
<td>3c</td>
<td>2.705</td>
<td>377.263</td>
<td>123.65</td>
<td>0.246</td>
<td>0.327</td>
<td>4.81</td>
</tr>
<tr>
<td>5c</td>
<td>0.628</td>
<td>446.773</td>
<td>123.65</td>
<td>0.216</td>
<td>0.276</td>
<td>4.56</td>
</tr>
<tr>
<td>5d</td>
<td>0.741</td>
<td>456.489</td>
<td>123.65</td>
<td>0.213</td>
<td>0.271</td>
<td>4.49</td>
</tr>
<tr>
<td>3o</td>
<td>2.668</td>
<td>386.493</td>
<td>112.22</td>
<td>0.225</td>
<td>0.290</td>
<td>4.29</td>
</tr>
<tr>
<td>3l</td>
<td>2.057</td>
<td>359.925</td>
<td>149.37</td>
<td>0.293</td>
<td>0.415</td>
<td>3.81</td>
</tr>
<tr>
<td>5a</td>
<td>0.013</td>
<td>421.489</td>
<td>112.22</td>
<td>0.210</td>
<td>0.266</td>
<td>3.75</td>
</tr>
</tbody>
</table>

\textsuperscript{a}logP\textsubscript{c-hex}: log partition between cyclohexane and water; \textsuperscript{b}HSA: hydrophobic surface area; \textsuperscript{c}PSA: polar surface areas; \textsuperscript{d}PSAR: polar surface and surface; \textsuperscript{e}PHSAR: polar hydrophobic surface area; \textsuperscript{f}pLC\textsubscript{50}: experimental activity.
Thus, we conclude that the substituent group on $R_1$ being activator of the aromatic ring is better for the activity, but this electro donor group needs to have an ideal hydrophilic character not to make the molecule with large polar surface area.

Conclusions

Compounds 3c and 3d, showed excellent activity as larvicides. In the evaluation of larvicidal activity and mammalian cell toxicity, pyrimidines which showed the best results, i.e., low toxicity in macrophages and high mortality rate at the concentration of 100 ppm for A. aegypti larvae, were: 3f, 3j, 3k and 5a. The QSAR models showed some physicochemical properties related to the pharmacokinetic behavior, such as hydrophobicity when the compounds were tested against mosquito larvae. A simple pattern of substituent groups, such as methyl and methoxy at $R_1$, and a succinimide at $R_3$, are responsible for the increase of larvicidal activity of the pyrimidine derivatives. This information can be used in further studies.

Supplementary Information

Supplementary data are available free of charge at http://jbcs.sbq.org.br as PDF file.

Acknowledgments

The authors gratefully acknowledge the Brazilian agencies FACEPE, CNPq, and CAPES, for providing financial support under grant: FACEPE/BFP-0094-4.03/17, Z. S. do Monte thanks to the Postgraduate Program in Pharmaceutical Sciences and, Postgraduate Program in Biological Sciences Chemistry Graduate Program at UFPE and, FACEPE and, CNPq/INCT-INAMI fellowships.

Author Contributions

Zenaide S. Monte was responsible for investigation (lead); Daniela M. A. F. Navarro for conceptualization and investigation (lead); Júlio C. R. O. F. Aguiar for investigation (supporting); Jessica S. Nascimento for investigation (supporting); Marcus T. Scorti for validation (lead); Luciana Scotti for formal analysis (lead); Renata P. C. Barros for investigation (supporting); Aline C. S. Santos for investigation (lead); Valéria R. A. Pereira for conceptualization (lead); Emerson P. S. Falcão for conceptualization (lead); Sebastião. J. Melo for project administration (lead).

References

27. GraphPad Prism, version 5.00 for Windows; GraphPad Software, San Diego, California, USA, 2015. Available at www.graphpad.com, accessed in February 2020.

Submitted: November 22, 2019
Published online: March 11, 2020