QSAR APPROACH TO ESTIMATING THE ANALGESIC ACTIVITY OF A SERIES OF TRI-SUBSTITUTED PYRIMIDINE DERIVATIVES

ABSTRACT

In the search for new molecules with improved analgesic activity, a series of twenty tri-substituted pyrimidine derivatives were subjected to a QSAR study. This study was conducted using Pearson Correlation Coefficients, Principal Component Analysis (PCA), Hierarchical Ascending Classification (HAC) and Multiple Linear Regression. The model obtained correlates the analgesic activity expressed by logAA, the energy of the highest occupied molecular orbital (E_HOMO), the dipole moment (\mu_D) and the lipophilic coefficient ACD/logP. Statistical indicators R² = 91.55%; RMSE = 0.2592; F = 39.720; Q^2_LOO = 0.858 and R^2_p =0.765, establish the internal performance and stability of the model. While its external predictive ability has been proved by the statistical parameters Q^2_pred = 0.727 and RMSEP = 0.403. The domain of applicability of the model developed, determined by the technique of levers and standardized residues, did not present any aberrant observations. The model thus established made it possible to determine the structures of eleven new tri-substituted pyrimidine derivatives having a better analgesic activity than the study molecules, at the dose of 10 mg/kg of body weight.

Keywords: Analgesic Activity, Quantum Chemistry, Pyrimidine Derivatives, QSAR Study

INTRODUCTION

The International Association for the Study of Pain (IASP) defines pain as "an unpleasant sensation and an emotional experience in response to actual or potential tissue damage or described in these terms" [1]. It has several components including an affective and emotional component and a behavioral component. The pain changes the behavior. It can then result in the loss of jobs and even the rejection of the patient by relatives. Thus pain has a very high social cost and economic cost [2-4]. It has become a public health and ethical problem [5, 6], which mobilizes policy makers through care systems [7, 8]. The fight against pain is therefore necessary, obligatory and permanent. Several varieties of drugs exist for this purpose such as aspirin, paracetamol, codeine and morphine as well as many nonsteroidal anti-inflammatory drugs (NSAIDs). But many of these varieties have harmful side effects for the body [9]. In the case of NSAIDs,
for example, the main adverse reactions commonly seen are nausea, vomiting, drowsiness, constipation, respiratory depression, gastrointestinal disturbances (mucosal irritation or perforation, ulcer, bleeding), and renal failure (renal failure) [10]. The design of new analgesics seeks to reduce side effects and enhance their effectiveness. This led to the synthesis of selective inhibitors of cyclooxygenase [11-13]. But research continues every year, around the world, many molecules [14-16] and other natural substances are tested for their analgesic properties [17-20]. Pyrimidine derivatives have a good place in these studies. Indeed, many natural substances [21, 22] and synthetic products include this heterocycle with six vertices. They possess several interesting biological properties such as antimycobacterial [23], cardioprotective [24], antimalarial [25], antiparasitic [26] and, of course, analgesic and anti-inflammatory properties [27-31]. Pyrimidine derivatives are thus synthesized and analyzed for their analgesic properties around the world [12, 27-31]. In the classical method, researchers use the SAR methodology (Relation Structure Activity) [32, 33], in order to synthesize the molecules likely to have the best effect. The discovery of the analgesic activity of the compounds thus synthesized is done in vivo, using rats or mice, using a certain number of accepted techniques (tail-flick, hot-plate, writhing) [34]. This approach is expensive in synthetic products and time, so it is not without damage for the animals used. Alternative methods are then scanned. This is why the QSAR methodology, which has emerged for some decades [33, 35], and which links biological activity quantitatively to the molecular structure, has entered this research process [36-40]. It should be noted, however, that for the analgesic activities of pyrimidine derivatives, a very limited number use this methodology [31]. The aim of this work is to develop a QSAR model, by multiple linear regression, able to predict the analgesic activity of pyrimidine derivatives from physicochemical.

2. MATERIAL AND CALCULATION METHODS

2.1 Analgesic Activity of Derivatives

The experimental database consists of a set of tri-substituted pyrimidine derivatives having central analgesic activity. These data come from studies by Vishal et al. [41] (9 derivatives) and Rajendra et al. [42] (11 derivatives) (Figure 1). The biological activity of these molecules was determined according to the same protocol, using the "Tail-Flick" technique [43, 44] which uses hot water at 55 °C ± 0.5. The percentage inhibition (PI), pain (Table 1), were recorded at the time of 120 minutes after the oral intake of the drug candidate, by the animal (mouse), at the body doses D of 100 mg / kg [41] and 10mg / kg [42]. The inhibition percentages were converted to logAA, following expression 01, below (Table 1). This transformation makes the experimental values unbounded and takes into account both the D dose and the substance administered [45], through its molar mass M.

\[
\log AA = \log \left( \frac{PI}{100 - PI} \right) - \log \frac{D}{M}
\]

PI is expressed in%, D in mg per kg of body weight and M in g / mol.

![Figure 1: General structure of pyrimidine derivatives](image)

2.2 Molecular Descriptors

The three descriptors of free-molecule selected for this study were calculated with the help of the Gaussian 03 [46] program, using the DFT and hybrid functional method B3LYP, used from the 6-31G (d,p) database and the ACD/ChemSketch program [47]. Gradient-corrected functionalities and hybrid functionalities such as B3LYP give better energies and are in good agreement with high-level ab initio methods [48]. The split-valence and double-dzeta bases (6-31G (d,p)) which is sufficiently extended and the fact of taking into account the polarization functions is important because it takes into account the free doublets of heteroatoms. These are the energy of the highest occupied molecular orbital (HOMO), the di-
pole moment $\mu_D$ and the octanol/water partition coefficient ACD/LogP. The energy of the HOMO is an energetic parameter of global molecular reactivity \[49\]. This energy expresses the susceptibility of the molecule to be attacked by electrophilic reagents. Because the highest occupied molecular orbital houses the last electrons of the molecule. Large values of $E_{\text{HOMO}}$ (negative values) favor the electrophilic reactions of the molecule: this one easily releases its last electrons. Thus it has a higher nucleophilic power. The dipolar moment $\mu_D$ is related to the global distribution of the negative and positive charges of the molecule, it reflects the polarity of this one. Large values of the dipole moment mark the importance of the polar character of a molecule \[50\]. The octanol/water partition coefficient ACD/LogP is estimated according to Pettauskas et al. from atomic contributions and structural fragments taking into account intramolecular interactions \[51\]. This descriptor is a parameter of transport and passage through cell membranes. Large logP values are indicative of high solubility in lipids and good penetration into cell membranes but this implies low aqueous solubility. This can disadvantage the transport of the drug by the blood plasma. The descriptor values for the twenty derivatives are collated in Table 1.

2.3 Statistical Analyses

2.3.1 Data Analysis

Principal Component Analysis (PCA) \[52\] and Hierarchical Ascending Classification (HAC) \[53, 54\], implemented in XLSAT software \[55\], were employed.

PCA is a method of data analysis using projection. It is mainly applied to reduce the number of explanatory variables. It also produces a visualization of the observations in a space with two or three dimensions, in order to identify homogeneous groups or on the contrary atypical observations. It is for this purpose that this method is used in this work.

HAC is also a method of data analysis. It operates groupings into homogeneous classes of individuals in a set, using a similarity or dissimilarity measurement criterion and a grouping strategy based on an aggregation criterion. Successive groupings of individuals according to the aggregation criterion produce the dendrogram which is a hierarchy of partitions.

2.3.2 Construction and Estimation of the QSAR Model

Multiple Linear Regression (MLR) in the XLSTAT program is used to generate structural-activity analgesic models using the training set. This method establishes a linear relationship between logAA, expressing the biological activity, and representing the magnitude to be explained, and the three independent descriptors, which constitute the explanatory variables. The constructed models are evaluated by the values of the statistical indicators such as the coefficient of determination $R^2$, the standard error RMCE and the value of the test of Fisher F. These statistical indicators are defined by the expressions below \[56, 57\]:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_{i,\text{exp}} - Y_{i,\text{pred}})^2}{\sum_{i=1}^{n}(Y_{i,\text{exp}} - \overline{Y}_{TS})^2}$$

$$\text{RMCE} = \sqrt{\frac{\sum_{i=1}^{n}(Y_{i,\text{exp}} - Y_{i,\text{pred}})^2}{n - p - 1}}$$

$$F = \frac{n - p - 1}{n} \frac{\sum(Y_{i,\text{exp}} - \overline{Y}_{TS})^2}{\sum(Y_{i,\text{exp}} - Y_{i,\text{pred}})^2}$$

In these expressions: $n$ is the number of molecules in the Training Set (TS), $p$ is the number of descriptors in the model. $Y_{i,\text{exp}}$ and $Y_{i,\text{pred}}$ are the experimental and predicted values of biological activity for molecule $i$; $\overline{Y}_{TS}$ is its average value of biological activity for the training set.

2.3.3 Internal Prediction and Model Stability

To test the internal predictive character of the model developed, the leave-one-out (LOO) cross validation technique is used. This technique involves removing an observation from the formation set, then constructing a new model and finally predicting, by this model,
the activity of the observation removed. The cross validation parameter $Q^2_{LOO}$, according to the expression 05 below, is calculated at the end of the cycle.

$$Q^2_{LOO} = 1 - \frac{\sum(Y_{\text{exp}} - Y_{\text{pred}})^2}{\sum(Y_{\text{exp}} - \bar{Y}_S)^2}$$  

The stability of the model is established by the randomization test of the dependent variable. To this end, by maintaining the matrix of independent descriptors, intact [58], the values of the analgesic activity are distributed, at random, to the compounds of the training set. After each distribution, a new model is constructed and its coefficient of determination $R^2_{\text{TR}}$ is raised. Finally, the validation parameter of Roy et al., $R^2_{v}$ is calculated according to expression 06 below [59, 60], after each randomization.

$$R^2_v = R^2 \cdot \sqrt{R^2 - R^2_{\text{TR}}}$$  

2.3.4 External Validation of the Model

As for the external predictive character of the model, it is established on the one hand by the prediction correlation coefficient $Q^2_{\text{pred}}$ and the predictive average prediction error (RMSEP), calculated according to the expressions 07 and 08 below:

$$Q^2_{\text{pred}} = 1 - \frac{\sum_{i=1}^{n_{\text{ext}}}(Y_{\text{exp}} - Y_{\text{pred}})^2}{\sum(Y_{\text{exp}} - \bar{Y}_A)^2}$$  

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n_{\text{ext}}}(Y_{\text{exp}} - Y_{\text{pred}})^2}{n_{\text{ext}}}}$$

$n_{\text{ext}}$ is the number of molecules of the test set, $Y_{\text{exp}}$ and $Y_{\text{pred}}$ are the experimental and predicted values of the dependent variable for the molecule $i$ belonging to the test set.

And on the other hand, by the validation criteria of Tropsha and Golbraikh [61], listed below:

Criterion 1: $Q^2_{\text{LOO}} > 0.5$; Criterion 2: $Q^2_{\text{pred}} > 0.6$; Criterion 3: $\frac{R^2 - R^2_{0}}{R^2} < 0.1$ with $0.85 < k < 1.15$;
Criterion 4: $\frac{R^2 - R^2_{0}}{R^2} < 0.1$ with $0.85 < k < 1.15$; Criterion 5: $|R^2 - R^2_{0}| < 0.3$;

$R^2$ Coefficient of correlation between predicted and observed activities with constant. $R^2_0$ Coefficient of correlation without constant, $k$ slope of the correlation line of $Y_{\text{pred}}$ as a function of $Y_{\text{exp}}$ passing through the origin. $R^2_{0}'$ Coefficient of correlation without constant, $k'$ slope of the correlation line of $Y_{\text{exp}}$ as a function of $Y_{\text{pred}}$ passing through the origin. It is appropriate that these five criteria are satisfied [62] to validate a model.
3. RESULTS AND DISCUSSION

3.1 Constitution of Training and Validation Sets

Below are some results of the data analysis:

Table 2: Pearson inter-correlation matrix (n) of the 3 selected descriptors and analgesic activity

<table>
<thead>
<tr>
<th>Variables</th>
<th>logAA</th>
<th>$\mu_D$</th>
<th>$E_{\text{HOMO}}$</th>
<th>ACD/LogP</th>
</tr>
</thead>
<tbody>
<tr>
<td>logAA</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_D$</td>
<td>0.0703</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$E_{\text{HOMO}}$</td>
<td>0.8514</td>
<td>-0.3091</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ACD/LogP</td>
<td>-0.1180</td>
<td>-0.3173</td>
<td>0.1580</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 2: Representation of the 20 compounds in the factorial space (F1, F2) of the PCA, comprising 81.58% of the total information.

Figure 3: Dendrogram of the partition hierarchy of the 20 derivatives with truncation line at dissimilarity level 5.67 for 5 homogeneous classes.

Table 2 and Figure 4 show the linear independence of the three descriptors and the strong correlation of the $E_{\text{HOMO}}$ descriptor with the biological activity.

The initial set of twenty compounds was subdivided into a training set (15) and a validation set (5), in the proportions 75% and 25% [63, 64], respectively using PCA and HAC. The PCA produced a visualization of the 20 observations in the factorial plane (F1, F2), representing 81.58% of the total information (figure 2). This visualization reveals four groups and an isolated observation which is DP17. As for the HAC, it has grouped into four homogeneous classes and an isolated observation all the observations, according to the dendrogram (Figure 3). To achieve this, Euclidean distances between observations, in the space defined by the descriptors, as dissimilarity criteria and Ward's method [65] as aggregation criteria were used. The four groups of the PCA as well as the four classes of the HAC contain the same observations. The five test-set derivatives (in bold italic) were selected by selecting one observation per group and the isolated observation DP17. The compounds of this set thus cover the whole of the dispersion and make it possible to take into account all the characteristics of the overall set [66].
3.2 Equation of the QSAR Model \( \mu \)

Equation of the model (Equation 9) establishes a linear dependence between the analgesic activity expressed by logAA and the three descriptors \( H_D \), \( E_{HOMO} \) and ACD/LogP. The positive sign of the coefficients of descriptors \( H_D \) and \( E_{HOMO} \) shows that the large values of these descriptors favor central analgesic activity. On the other hand, the negative sign of the coefficient of ACD/LogP indicates that logAA and ACD/LogP evolve in opposite directions. High values of this third descriptor do not promote analgesic activity.

\[
\text{logAA}_{1}^{\text{Pred}} = 11.342(\pm 0.951)+0.234(\pm 0.078)^{H_D}+1.829(\pm 0.178)^{E_{HOMO}}-0.112(\pm 0.110)^{ACD/\text{LogP}} \quad (09)
\]

\( N = 15, \quad R^2 = 91.55\%; \quad \text{RMCE} = 0.2592; \quad F = 39.7201 \text{ and } p \quad 0.0001. \)

The value of the coefficient of determination \( R^2 \) indicates a model capable of explaining 91.55\% of the experimental variance; the values estimated by the model are close to the experimental values. This result is corroborated by the relatively small value of the mean squared difference \( \text{RMCE} = 0.2592 \). As for the Fisher test, its \( F = 39.7201 \) value confirms the significance of the model developed, in that it is greater than the required critical value \( F(11; 15; 0.05) = 2.507 \). Experimental and predicted values and residues are summarized in Table 1.

Contribution of descriptors to the model

According to the absolute values of the normalized coefficients (Figure 4), the importance of the weight of the descriptors involved in the model (Equation 9) decreases in the following order: \( E_{HOMO} > H_D > ACD/\text{LogP} \). The HOMO energy has the largest contribution with 0.9767 and the smallest ACD/LogP with (-0.1046), while \( \mu_D \) has 0.3253 as the contribution. The analgesic activity of these derivatives is therefore favored by molecules having a high nucleophilic character, a high polarity and low values of the lipophilic coefficient. Any proposal for a molecule to improve the central analgesic activity of these derivatives should take into account these criteria.

3.3 Internal and External Validations

The leave-one-out (LOO) cross-validation and randomization techniques of the explained variable were applied on the fifteen compounds of the training set. The statistical parameters of these validations are:

\[
Q^2_{\text{LOO}} = 0.8576 \quad \text{and} \quad R^2_p = 0.7616 \quad \text{after 10 randomizations [68].}
\]

The cross-correlation coefficient \( Q^2_{\text{LOO}} > 0.5 \) which is the minimum required value, according to Tropsha et al. [69]. The model developed therefore has a satisfactory internal predictive character. Also, the average value of the Roy et al. Parameter, \( R^2_p \), obtained after the ten randomizations is also greater than the minimum required value which is also 0.5 as well. This shows that the constructed model is not the result of chance [59-66]. All of these statistical parameters indicate that the developed model has a satisfactory internal predictive character and can be considered robust [56] and stable.

According to the external prediction parameters, the elaborated model is able to explain the activity of the unknown compounds of the test set with a prediction correlation coefficient \( Q^2_{\text{pred}} = 0.7271 \) and a Predictive Mean Prediction (RMSEP) error of 0.4033. This value is less than the standard deviation of the experimental values of the formation set which is \( \sigma = 0.7905 \).

Concerning the validation criteria of Tropsha and Golbraikh, we have:

Criterion 1: \( Q^2_{\text{LOO}} = 0.8576 > 0.5 \); Criterion 2: \( Q^2_{\text{pred}} = 0.7271 > 0.6 \); Criterion 3: \( \frac{R^2_{\text{LOO}}}{R^2} = 0.002 < 0.1 \) with \( 0.85 \leq k \leq 1.15 \); Criterion 4: \( \frac{R^2_{\text{LOO}}}{R^2} = 0.026 < 0.1 \) with \( 0.85 \leq k \leq 1.1341 \leq 1.15 \);
Criterion 5: $|R_{\text{test}}^2 - R_{\text{train}}^2| = 0.002 < 0.3$

These different values make it possible to affirm that the model developed enjoys a good external predictive performance.

### 3.4 Applicability Domain

The applicability domain of the model was defined using the standardized residue technique represented as a function of the levers $h_{ii}$ [70], calculated according to the expression 10 below.

$$h_{ii} = x_i (X^T X)^{-1} x_i^T$$

Where $x_i$ is the line vector of the descriptors of the compound $i$ and $X$ is the matrix built on the values of the descriptors of the model and the compounds of the training set. The exponent $t$ refers to the matrix or transposed vector. The MINITAB software was used to calculate the levers of the molecules of the training set as well as those of the molecules of the test set. Figure 6 below shows the graph obtained. For the $N = 15$ compounds of the formation set and the $k = 3$ descriptors of the model, the threshold value of the levers $h^* = 3 \frac{k+1}{N}$ is 0.800. The standard deviation of the experimental values of the training set is $\sigma = 0.7905$. The extreme values of the residues are $\pm 3\sigma$, i.e., -2.372 and +2.372. These different values delimit the domain of applicability of the model [70-73].

The graph (Figure 6) shows that all the observations of the training and validation sets belong to the defined applicability domain. No aberrant observations can be reported for the model developed [72]. Thus, for the model developed to predict the analgesic activity of a new molecule, its lever $h_{ii}$ must be below the threshold value $h^* = 0.800$. This molecule must belong to the domain of applicability of the model [74].

The experimental and predicted values as well as the residues obtained for the compounds of the two sets are given in Table 1. Figures 7 and 8 show the regression line of the predicted and experimental values as well as the graph of similarity.

**Figure 5**: Standardized coefficients of descriptors of the developed model.

**Figure 6**: William’s plot of standardized residues according to the leverages.

**Figure 7**: Graph of Predicted Values Based on Experimental Values.

**Figure 8**: Graph of experimental values versus logAA predicted values.
Table 1: Designation codes, Natures of substituents and percentages of pain inhibition after 120 minutes of derivation of the derivative. Values of the three model descriptors, experimental and predicted values and calculated residues for each compound of the training and validation set.

<table>
<thead>
<tr>
<th>Compounds</th>
<th>Ar</th>
<th>R</th>
<th>YHn</th>
<th>( E_{\text{HOMO}} ) (eV)</th>
<th>ACD/LogP</th>
<th>PI(%)</th>
<th>Exp(logAA)</th>
<th>Pred(logAA)</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>DP01</td>
<td>C(_6)H(_5)</td>
<td>BrC(_6)H(_4)</td>
<td>OH</td>
<td>4.252</td>
<td>-6.208</td>
<td>2.440</td>
<td>57.03</td>
<td>0.623</td>
<td>0.713</td>
</tr>
<tr>
<td>DP02</td>
<td>C(_6)H(_5)</td>
<td>BrC(_6)H(_4)</td>
<td>OH</td>
<td>4.278</td>
<td>-6.227</td>
<td>2.790</td>
<td>61.97</td>
<td>0.733</td>
<td>0.645</td>
</tr>
<tr>
<td>DP04</td>
<td>C(_6)H(_5)</td>
<td>ClC(_6)H(_4)</td>
<td>SH</td>
<td>4.321</td>
<td>-6.216</td>
<td>2.100</td>
<td>74.90</td>
<td>0.909</td>
<td>0.751</td>
</tr>
<tr>
<td>DP06</td>
<td>C(_6)H(_5)</td>
<td>ClC(_6)H(_4)</td>
<td>SH</td>
<td>2.430</td>
<td>-5.885</td>
<td>3.640</td>
<td>59.69</td>
<td>0.604</td>
<td>0.744</td>
</tr>
<tr>
<td>DP07</td>
<td>C(_6)H(_5)</td>
<td>FC(_6)H(_4)</td>
<td>NH(_2)</td>
<td>3.928</td>
<td>-6.154</td>
<td>1.470</td>
<td>69.20</td>
<td>0.760</td>
<td>0.844</td>
</tr>
<tr>
<td>DP08</td>
<td>C(_6)H(_5)</td>
<td>FC(_6)H(_4)</td>
<td>NH(_2)</td>
<td>4.092</td>
<td>-6.175</td>
<td>1.820</td>
<td>72.24</td>
<td>0.850</td>
<td>0.805</td>
</tr>
<tr>
<td>DP09</td>
<td>C(_6)H(_5)</td>
<td>FC(_6)H(_4)</td>
<td>NH(_2)</td>
<td>1.791</td>
<td>-5.820</td>
<td>3.010</td>
<td>56.27</td>
<td>0.516</td>
<td>0.783</td>
</tr>
<tr>
<td>DP11</td>
<td>C(_6)H(_5)</td>
<td>C(_6)H(_5)</td>
<td>NH(_2)</td>
<td>4.018</td>
<td>-5.428</td>
<td>2.610</td>
<td>90.23</td>
<td>2.413</td>
<td>2.065</td>
</tr>
<tr>
<td>DP13</td>
<td>C(_6)H(_5)</td>
<td>C(_6)H(_5)</td>
<td>NH(_2)</td>
<td>2.554</td>
<td>-5.358</td>
<td>2.160</td>
<td>90.45</td>
<td>2.484</td>
<td>1.901</td>
</tr>
<tr>
<td>DP14</td>
<td>C(_6)H(_5)</td>
<td>C(_6)H(_5)</td>
<td>NH(_2)</td>
<td>3.554</td>
<td>-5.342</td>
<td>1.710</td>
<td>78.46</td>
<td>2.108</td>
<td>2.214</td>
</tr>
<tr>
<td>DP15</td>
<td>C(_6)H(_5)</td>
<td>C(_6)H(_5)</td>
<td>NH(_2)</td>
<td>2.739</td>
<td>-5.346</td>
<td>3.130</td>
<td>75.37</td>
<td>1.927</td>
<td>1.859</td>
</tr>
<tr>
<td>DP16</td>
<td>C(_6)H(_5)</td>
<td>C(_6)H(_5)</td>
<td>NH(_2)</td>
<td>2.982</td>
<td>-4.966</td>
<td>2.780</td>
<td>88.63</td>
<td>2.376</td>
<td>2.649</td>
</tr>
<tr>
<td>DP18</td>
<td>C(_6)H(_5)</td>
<td>2-C(_6)H(_5)</td>
<td>NH(_2)</td>
<td>1.958</td>
<td>-5.310</td>
<td>1.330</td>
<td>75.41</td>
<td>1.907</td>
<td>1.942</td>
</tr>
<tr>
<td>DP19</td>
<td>C(_6)H(_5)</td>
<td>4-C(_6)H(_5)</td>
<td>NH(_2)</td>
<td>5.441</td>
<td>-5.529</td>
<td>1.290</td>
<td>87.22</td>
<td>2.254</td>
<td>2.361</td>
</tr>
<tr>
<td>DP20</td>
<td>C(_6)H(_5)</td>
<td>3-C(_6)H(_5)</td>
<td>NH(_2)</td>
<td>5.040</td>
<td>-5.478</td>
<td>1.360</td>
<td>84.79</td>
<td>2.166</td>
<td>2.353</td>
</tr>
</tbody>
</table>

The absolute values of the residues for both sets are all smaller than the standard deviation of the training set values \( \sigma = 0.7905 \).
3.5 Improvement of Analgesic Activity

In order to obtain derivatives having an analgesic activity higher than the maximum experimental activity, structural modifications were made to the molecules having the best experimental activities, namely: DP10 with 2.546; DP13 with 2.484; DP11 with 2.413 and DP16 with 2.376. Sixteen (16) tri-substituted derivatives were then proposed. After optimization, then calculation of the frequencies at the level B3LYP/6-31G (d, p), the dipole moment as well as the energy of the HOMO of each of the sixteen proposed molecules were recorded. The coefficient of lipophilicity ACD/logP was calculated for each of the molecules. Equation 9 was used to calculate logAA for each proposition. Their levers were also calculated. Table 3 contains the designation codes, the 2D structures, the values of the three descriptors, those of logAA predicted as well as the levers of the proposed molecules.

Of the 16 proposals, 12 have a logAA value greater than the maximum experimental value. A proposition has a leverage greater than the threshold, of value 0.937. This is P09 which is part of the twelve. This outlier could be due to the high value of the dipole moment, ie 6.497, which is the highest. Thus the model was able to find the structures of eleven new molecules similar to the basic molecules used and having a higher central analgesic activity. It is P02, P05, P06, P07, P08, P10, P11, P13, P14, P15 and P16.

Table 3: Designation codes, 2D structures, values of the three descriptors and logAA predicted by the model (Equation 9) as well as the $h_{ij}$ levers of the proposed molecules

<table>
<thead>
<tr>
<th>Code</th>
<th>Structure 2D</th>
<th>$\mu_D$ (D)</th>
<th>$E_{\text{HOMO}}$ (eV)</th>
<th>ACD/LogP</th>
<th>Préd (logAA)</th>
<th>$h_{ij}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>P01</td>
<td><img src="image" alt="Structure P01" /></td>
<td>6.404</td>
<td>-5.700</td>
<td>2.05</td>
<td>2.400</td>
<td>0.703</td>
</tr>
<tr>
<td>P02</td>
<td><img src="image" alt="Structure P02" /></td>
<td>5.000</td>
<td>-5.347</td>
<td>0.62</td>
<td>2.846</td>
<td>0.617</td>
</tr>
<tr>
<td>P03</td>
<td><img src="image" alt="Structure P03" /></td>
<td>5.089</td>
<td>-5.586</td>
<td>0.97</td>
<td>2.380</td>
<td>0.447</td>
</tr>
<tr>
<td>P04</td>
<td><img src="image" alt="Structure P04" /></td>
<td>5.901</td>
<td>-5.620</td>
<td>1.07</td>
<td>2.539</td>
<td>0.649</td>
</tr>
<tr>
<td>P05</td>
<td><img src="image" alt="Structure P05" /></td>
<td>4.997</td>
<td>-5.056</td>
<td>1.24</td>
<td>3.297</td>
<td>0.635</td>
</tr>
<tr>
<td>P06</td>
<td><img src="image" alt="Structure P06" /></td>
<td>5.681</td>
<td>-5.134</td>
<td>1.59</td>
<td>3.303</td>
<td>0.735</td>
</tr>
<tr>
<td>P07</td>
<td><img src="image" alt="Structure P07" /></td>
<td>6.044</td>
<td>-5.216</td>
<td>2.60</td>
<td>3.115</td>
<td>0.758</td>
</tr>
<tr>
<td>P08</td>
<td><img src="image" alt="Structure P08" /></td>
<td>5.650</td>
<td>-5.320</td>
<td>1.06</td>
<td>3.024</td>
<td>0.704</td>
</tr>
</tbody>
</table>
4. CONCLUSION

This work has shown that it is possible to explain the central analgesic activity of tri-substituted pyrimidine derivatives by the $\mu_D$, $E_{\text{HOMO}}$ and $\text{ACD/LogP}$ descriptors. Both quantum descriptors were calculated by the DFT method at the B3LYP/6-31 (d, p) theory level. Multiple linear regression (MLR) has been used to develop a QSAR model. The energy of the highest occupied molecular orbital $E_{\text{HOMO}}$ has, in this model, the largest contribution followed by the dipolar momentum $\mu_D$ and finally the coefficient of lipophilia $\text{ACD/LogP}$. The established model can be considered reliable, robust and efficient capable of predicting the analgesic activity of new pyrimidine compounds belonging to the defined field of applicability, given the values of the different statistical estimators ($R^2 = 91.55\%$, $RMCE = 0.2592$, $F = 39.7201$, $Q^2_{\text{loo}} = 85.76$, $Q^2_{\text{pred}} = 72.71\%$). Thus, this model made it possible to determine the structures of eleven new tri-substituted derivatives of pyrimidine having a better analgesic activity than that of the starting molecules. A synthesis of these molecules should be performed to verify these theoretical predictions.


[6] Direction de l’Hospitalisation et de l’Organisation des Soins, Ministère de la Santé, de la Famille et des Personnes Handicapées (France), Lutter Contre la Douleur, Guide d’Orientation, Organiser la Lutte Contre la Douleur dans les Etablissements de Santé, 0AA68d01


[33] H. V. De Waterbeemd and S. Rose, Quantitative Approaches to Structure-Activity Relat-


[64] Z. Hajimahdi, F. Safizadeh and A. Zarghi., QSAR Analysis for Some 1, 2-Benzisothiazol


[71] M. Ghamali, S. Chitta, M. Bouachrine and T. Lakhlihi, Méthodologie générale d’une étude RQSA/RQSP, Revue Interdisciplinaire, 2016, 1(1), 1-6

