Quantum and Structural Molecular Fragment models used to predict anti-inflammatory activity

ABSTRACT
In this paper, we predict the anti-inflammatory activity of a series of 26 structures of N-arylanthranilic acid. So, Quantitative Structure-Activity Relationship (QSAR) method remains the focus of many studies aimed at modeling and prediction of physicochemical properties or biological activities of molecule. Two models were used: quantum model and Structural Molecular Fragment (SMF) model. In the first model, semi-empirical (AM1) approach was used to calculate the quantum chemical descriptors using GAUSSIAN 09 package and the others chemical descriptors were calculated with chemaxon package. In the second model, Structural Molecular Fragment were generated by I.S.I.D.A (In Silico Design and Data Analysis). Our two models were built by using a Multiple Linear Regression Analysis (MLR). The concluded QSAR models reflected that the drugs activity was mainly attributed to quantum chemical descriptors with the statistical analysis of multiple R-squared equal to 0.9898 v. s 0.9077 for the Structural Molecular Fragment developed in I.S.I.D.A.

Keywords: N-arylanthranilic acids, anti-inflammatory activity, quantum descriptors, Structural Molecular Fragment.
1. INTRODUCTION

Non-Steroidal Anti-Inflammatory Drugs (NSAIDS) are still the most prescribed drugs worldwide for the treatment of inflammatory diseases like rheumatoid arthritis, osteoarthritis, orthopedic injuries, post-operative pain, acute myalgias etc. [1, 2]. N-aryl anthranilic acids belong to the category of NSAIDS. They are amino isosteres of salicylates and are also known as fenemates. Important molecules of this class include mefenamic acid, flufenamic acid and meclofenamic acid. Fenemates act by blocking the metabolism of arachidonic acid by the enzyme cyclooxygenase (COX), one of the key enzymes in the arachidonic acid cascade [3-5]. This enzyme bis-oxygenates arachidonic acid to prostaglandine G2, which is subsequently degraded to vasoactive and inflammatory mediators such as prostaglandins (PGS), prostacyclin (PGI2), and thromboxane-A2[6]. Some fenemates also inhibit arachidonic acid lipoxygenase resulting in decreased synthesis of leukotrienes, known mediators involved in inflammatory process [7]. Studies suggest that flufenamic and tolfenamic acids suppress proliferation of human peripheral blood lymphocytes by a mechanism, which involves inhibition of Ca\(^{2+}\) influx and is not related to inhibition of prostanoid synthesis [8].

Quantitative structure-activity and structure property relationship (QSAR/QSPR) studies are unquestionably of great importance in modern chemistry and biochemistry [9]. The concept of QSAR/QSPR is to transform searches for compounds with desired properties using chemical intuition and experience into a mathematically quantified and computerized form [10]. Once a correlation between structure and activity/property is found, any number of compounds, including those not yet synthesized, can be readily screened on the computer to select structures with the properties desired [11, 12]. It is then possible to select the most promising compounds to synthesize and test in the laboratory. Thus, the QSAR/QSPR approach conserves resource and accelerates the process of development of new molecules for use as drugs, materials, additives, or for any other purpose [13-16].

In the present study, relationship of chemical quantum and structural molecular fragment with anti-inflammatory activity of N-aryl anthranilic acids derivatives has been investigated and suitable models developed for the prediction of anti-inflammatory activity.

2. MATERIALS AND METHODS

In this study, we used the following materials: gaussi-an09 [17], chemaxon [18], ISIDA/QSPR [19-37], R [38] and data set of the 26 anthranilic acids molecule belonging to a group of NSAIDs were taken from the literature with their experimental activities (table.1) [39].

Figure 1. chemical structure of anthranilic acid

Table 1. A dataset of 26 N-aryl anthranilic acids with anti-inflammatory activity [22]

| Mol | R1  | R2  | R3  | R4  | R5  | MED\(^a\) | \(A_{exp}\) |
|-----|-----|-----|-----|-----|-----|---------|---------|
| 1   | Cl  | H   | CF3 | H   | Cl  | 0.8     | 3.699   |
| 2   | CH3 | SO\(_2\)N(CH\(_3\))\(_2\) | H   | H   | CH\(_3\) | 0.5    | 3.903   |
| 3   | CH\(_2\) | NH\(_2\) | H   | H   | Cl  | 6.2    | 2.809   |
| 4   | CH\(_2\) | CH\(_3\) | H   | H   | Cl  | 12.5   | 2.505   |
| 5   | Cl  | Cl  | H   | H   | CH\(_3\) | 0.8    | 3.699   |
| 6   | Cl  | H   | C\(_6\)H\(_5\) | H   | Cl  | 0.8    | 3.699   |
| 7   | Cl  | H   | Cl  | Cl  | H   | 400    | 1.080   |
| 8   | Cl  | Cl  | Cl  | H   | H   | 200    | 1.301   |
| 9   | Cl  | H   | Cl  | Cl  | H   | 100    | 1.602   |
| 10  | NH\(_2\) | CH\(_3\) | H   | H   | CH\(_3\) | 25     | 2.204   |
| 11  | CH\(_3\) | CH\(_3\) | H   | H   | CH\(_3\) | 6.2    | 2.809   |
| 12  | Cl  | CH\(_3\) | H   | H   | CH\(_3\) | 3.1    | 3.110   |
| 13  | CH\(_2\) | Cl  | H   | CH\(_3\) | H   | 1.6    | 3.397   |
| 14  | CH\(_2\) | C\(_6\)H\(_5\) | H   | H   | CH\(_3\) | 1.6    | 3.397   |
| 15  | CH\(_2\) | NH\(_2\) | H   | H   | Cl  | 1.3    | 3.488   |
| 16  | CH\(_3\) | SO\(_2\)CH\(_3\) | H   | H   | CH\(_3\) | 0.6    | 3.823   |
| 17  | Cl  | N(CH\(_3\))\(_2\) | H   | H   | Cl  | 0.6    | 3.823   |
| 18  | CH\(_2\) | SOCH\(_3\) | H   | H   | CH\(_3\) | 0.5    | 3.903   |
| 19  | Cl  | Cl  | Cl  | H   | CH\(_3\) | 12.5   | 2.505   |
| 20  | CH\(_2\) | CH\(_3\) | H   | CH\(_3\) | CH\(_3\) | 100    | 1.602   |
| 21  | Cl  | Cl  | Cl  | H   | Cl  | 12.5   | 2.505   |
| 22  | Cl  | CH\(_3\) | Cl  | H   | Cl  | 12.5   | 2.505   |
| 23  | Cl  | Cl  | Cl  | Cl  | H   | 100    | 1.602   |
| 24  | Cl  | Cl  | H   | Cl  | Cl  | 1.6    | 3.397   |
| 25  | Cl  | Cl  | Cl  | Cl  | Cl  | 25     | 2.204   |
| 26  | CH\(_2\) | CH\(_3\) | Cl  | CH\(_3\) | Cl  | 100    | 1.602   |

The biological activity \(A\) was calculated from the minimal effective dose (MED mg/kgbody) by formula: \(A = \log(4000/MED)\)
To perform our two models, we are using Multiple Linear Regression Analysis (MLR). The first model used quantum descriptors; 23 quantum chemical descriptors were computed with Gaussian 09. The semi-empirical AM1 method was employed for the calculation of these descriptors (Table 2). Chemaxon software with Marvin suite is a chemically intelligent desktop toolkit built to help us draw, edit, publish, render, import and export chemical structures and as well as allowing us to convert between various chemical and graphical file formats. Software R provides a wide variety of statistical (linear and nonlinear modeling, classical statistical test) and graphical techniques.

Heuristic method was applied to the whole dataset of the N-arylanthranilic acids, a pre-selection of descriptors occurs. Descriptors unavailable for some compounds are discarded altogether with the invariant descriptors and descriptors that correlate poorly. Additional descriptors are discarded when high inter-correlations between them are found. The remaining descriptors are then ranked according to their correlation coefficients.

Table 2. calculated quantum chemical descriptors of anthranilic acids (1-26)
Table 2: (continued) calculated quantum chemical descriptors of anthranilic acids (1-26)

|    | Q12Q | D13Q | D14Q | D15Q | D16Q | D17Q | D18Q | D19Q | D20Q | D21Q | D22Q | D23Q |
|----|------|------|------|------|------|------|------|------|------|------|------|------|
| 1  | 124.929 | -1.311 | -1.652 | -3.646 | 0.139 | 7.191 | 0.194 | -0.194 | 0.136 | 182.629 | 28473.499 | 17.749 |
| 2  | 166.065 | 2.787 | 0.618 | -8.791 | 0.138 | 7.263 | 0.190 | -0.190 | 0.131 | 218.036 | 13638.764 | 85.433 |
| 3  | 128.524 | -2.464 | 1.692 | -3.295 | 0.137 | 7.287 | 0.182 | -0.182 | 0.121 | 177.505 | 24775.831 | 19.794 |
| 4  | 129.211 | -1.933 | 0.44  | -3.151 | 0.136 | 7.331 | 0.181 | -0.181 | 0.120 | 178.591 | 23409.146 | 13.861 |
| 5  | 120.605 | 0.481 | 3.857 | -4.158 | 0.137 | 7.293 | 0.186 | -0.186 | 0.127 | 179.633 | 24219.272 | 32.399 |
| 6  | 130.877 | -2.456 | 1.549 | -3.366 | 0.137 | 7.283 | 0.183 | -0.183 | 0.122 | 189.417 | 29260.484 | 19.767 |
| 7  | 115.183 | -0.704 | -1.171 | -3.937 | 0.138 | 7.273 | 0.192 | -0.192 | 0.133 | 185.331 | 39489.304 | 17.372 |
| 8  | 112.254 | -1.46  | -1.457 | -5.062 | 0.137 | 7.282 | 0.191 | -0.191 | 0.133 | 184.683 | 39156.247 | 29.877 |
| 9  | 115.767 | 0.082 | -1.651 | -3.542 | 0.138 | 7.220 | 0.190 | -0.190 | 0.130 | 181.583 | 35663.125 | 15.280 |
| 10 | 120.36  | -0.48  | 0.251  | 4.518  | 0.134 | 7.485 | 0.181 | -0.181 | 0.122 | 177.501 | 23202.902 | 20.657 |
| 11 | 124.914 | 1.525  | 0.291  | -3.751 | 0.133 | 7.499 | 0.176 | -0.176 | 0.116 | 179.558 | 23593.282 | 16.484 |
| 12 | 126.315 | -1.464 | 2.502  | -3.992 | 0.136 | 7.362 | 0.180 | -0.180 | 0.119 | 178.933 | 21592.401 | 24.344 |
| 13 | 126.763 | -0.247 | 2.403  | -4.057 | 0.135 | 7.385 | 0.183 | -0.183 | 0.123 | 182.592 | 26309.491 | 22.297 |
| 14 | 135.063 | -1.466 | 0.621  | -3.773 | 0.133 | 7.504 | 0.176 | -0.176 | 0.116 | 188.361 | 24739.274 | 16.777 |
| 15 | 144.089 | -2.25  | 1.795  | -3.295 | 0.137 | 7.302 | 0.181 | -0.181 | 0.120 | 195.707 | 25321.224 | 19.149 |
| 16 | 157.461 | 4.881  | 1.649  | -8.439 | 0.130 | 7.699 | 0.200 | -0.200 | 0.154 | 214.409 | 22972.847 | 97.773 |
| 17 | 138.271 | -1.976 | 3.081  | -3.405 | 0.138 | 7.245 | 0.185 | -0.185 | 0.124 | 195.253 | 23344.798 | 25.000 |
| 18 | 136.104 | 0.2    | 4.833  | -2.259 | 0.135 | 7.421 | 0.184 | -0.184 | 0.126 | 208.641 | 28770.610 | 28.505 |
| 19 | 120.823 | 1.522  | 3.577  | -4.152 | 0.137 | 7.317 | 0.190 | -0.190 | 0.132 | 192.114 | 37174.421 | 32.353 |
| 20 | 141.262 | 2.054  | -1.694 | -3.376 | 0.133 | 7.524 | 0.175 | -0.175 | 0.115 | 189.621 | 25369.852 | 18.490 |
| 21 | 115.935 | 1.116  | 2.53   | -3.563 | 0.139 | 7.191 | 0.194 | -0.194 | 0.135 | 191.584 | 37231.460 | 20.349 |
| 22 | 122.702 | -0.3   | 1.347  | -3.409 | 0.138 | 7.241 | 0.189 | -0.189 | 0.129 | 190.418 | 35514.969 | 13.528 |
| 23 | 115.943 | -1.673 | -2.543 | -3.861 | 0.139 | 7.210 | 0.195 | -0.195 | 0.137 | 194.968 | 39995.788 | 24.177 |
| 24 | 116.463 | 0.062  | 1.88   | -3.401 | 0.139 | 7.184 | 0.193 | -0.193 | 0.134 | 189.496 | 25883.499 | 15.101 |
| 25 | 116.199 | -1.403 | -1.871 | -3.514 | 0.139 | 7.169 | 0.197 | -0.197 | 0.139 | 201.574 | 38642.089 | 17.825 |
| 26 | 135.721 | -0.465 | 0.563  | -3.298 | 0.135 | 7.383 | 0.183 | -0.183 | 0.123 | 200.052 | 36327.423 | 11.418 |

D1Q: charge max, D2Q: charge min, D3Q: HOMO-energy, D4Q: LUMO-energy, D5Q: thermal energy, D6Q: constant volume molar heat capacity, D7Q: entropy, D8Q: partition function, D9Q: molecular dipole moment, D10Q: polarizability- \( \alpha \), D11Q: polarizability- \( \alpha_n \), D12Q: polarizability- \( \alpha_s \), D13Q: component of dipole along inertia axe x, D14Q: component of dipole along inertia axe y, D15Q: component of dipole along inertia axe z, D16Q: absolute hardness, D17Q: inverse of hardness, D18Q: chemical potential, D19Q: electro negativity D20Q: electrophilicity index, D21Q: mean polarizability of molecule D22Q: anisotropy of polarisability, D23Q: square of molecular dipole moment.
To decide whether a model generated is good or not is commonly defined by the square coefficient of fitting model ($R^2$), adjusted R-squared ($R^2_{adj}$) and Fisher Statistic ($F$).

$$R^2 = \frac{SCE}{SCT} = \frac{\Sigma (A_i - \bar{A})^2}{\Sigma (A_i - \bar{A})^2}$$

$$R^2_{adj} = 1 - (1 - R^2) \frac{N-1}{N-k}$$

$$F = \frac{SCE / (N-1)}{SCT / (N-k)}$$

With $SCE = \Sigma (A_i - \bar{A})^2$ and $N$: the number of individuals, $k$: the variables

The second model used is Structural Molecular Fragment (SMF), this method is developed in ISIDA/QSPR, the latest is based on the splitting of a molecular graph on fragments (subgraphs), and on the calculation of their contributions to a given property $Y$. Two classes of fragments are used: “sequences” (I) and “augmented atoms” (II). Three sub-types of AB, A and B are defined for each class. For the fragments I, they represent sequences of atoms and bonds (AB), of atoms only (A), or of bonds only (B). Shortest or all paths from one atom to the other are used. For each type of sequences, the minimal ($n_{\text{min}}$) and maximal ($n_{\text{max}}$) number of constituted atoms must be defined. Thus, for the partitioning I (AB, $n_{\text{min}} - n_{\text{max}}$), I (A, $n_{\text{min}} - n_{\text{max}}$) and I (B, $n_{\text{min}} - n_{\text{max}}$), the program generates “intermediate” sequences involving $n$ atoms ($n_{\text{min}} \leq n \leq n_{\text{max}}$). In the current version of ISIDA/QSPR, $n_{\text{min}} \geq 2$ and $n_{\text{max}} \leq 15$. The number of sequence’s types of different length corresponding to $n_{\text{min}} = 2$ and $n_{\text{max}} = 15$ is equal to 105 for each of three subtypes AB, A and B, totally 315 types of sequences. QSPR modeling was performed using Multiple Linear Regression Analysis (MLR) of the ISIDA/QSPR program with combined forward and backward stepwise variable selection techniques. MLR is applied to build linear relationships between independent variables (SMF descriptors: $N_i$ $i = 1, 2...$) and a dependent variable (here target property $Y = A$)

$$A_{\text{cal}} = a_0 + \Sigma a_i N_i$$

where every descriptor value is associated with observed property value ($A$), $a_i$ is descriptor contribution, and $a_0$ is the independent term which is omitted in a part of models see table (3). The Singular Value Decomposition method is used to fit contributions $a_i$ and to minimize the sum of squared residuals which are squared differences between the property values calculated by the model ($A_{\text{cal}}$) and observed values $A_{\text{exp}}$ in the training set. The program can generate more than 25,000 MLR models; each of them corresponds to particular type of the SMF descriptors and MLR equation ($a_0 = 0$ or $a_0 \neq 0$) and applied variable selection technique.

### Table 3. Example of ISIDA Model

| Molecule | The contribution matrix $M_{ij}$ | $A_{\text{cal}}$ |
| --- | --- | --- |
| SMF(N): | $N_{c-c-c-c-c}$ | $N_{c-c-c-c}$ | $N_{c-o}$ | $N_{c-c}$ | $N_{c-c-o}$ |
| 1 | 0 | 1 | 10 | 5 | 0 | -0.222 |
| 2 | 0 | 8 | 1 | 4 | 0 | 0.973 |
| 3 | 0 | 4 | 1 | 2 | 4 | -0.056 |

\[ A_{\text{cal,1}} = -0.36 \times N_{c-c-c-c-c-o} - 0.29 \times N_{c-o} + 0.12 \times N_{c-c-c} \]
To validate consensus model, the external 5-fold cross validation (5-CV) was applied. [40-42] ISIDA, implicitly keeps every 5th compound in the test set, the initial set was randomly split into 5 subsets, each of which was iteratively ignored at the training stage, to serve as internal validation set while the four others formed, together, the learning set. For each of these 5 splitting schemes, models were built followed by prediction calculations on the corresponding validation set. Finally, all values calculated for five test sets are merged into one file to analyse overall linear correlations between experimental and predicted property. One can use Determination Coefficient (R²), Root Mean Squared Error (RMSE) or Mean Average Error (MAE), to estimate the quality of the linear correlation between predicted ($A_{pred}$) and experimental ($A_{exp}$) data for n compounds. Formulas for the statistical parameters are formulated below.

**Root–Mean Square Error**

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (A_{pred,i} - A_{exp,i})^2}$$

(5)

**Mean Average Error**

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |A_{pred,i} - A_{exp,i}|$$

(6)

ISIDA calculates a Consensus Model (CM) combining the information issued from several models. At the first step, hundreds of models are built using different initial pools of descriptors corresponding to different fragmentation types.

The contributions of $a_i$ are calculated by minimizing a functional

$$U(a_i) = \sum_{i=1}^{n} W_i (A_{exp,i} - A_{calc,i})^2$$

(7)

where n is the number of the compounds in the training set, wi the weight accounting for the accuracy of the experimental data, $A_{exp}$ and $A_{calc}$ are, respectively, experimental and calculated.

Linear equations (4) are obtained by the contribution matrix $M_{ij}$

$$M_{ij} = \text{molecule (row)} \times \text{fragment (column)}$$

(8)

### 3. RESULTS AND DISCUSSION

#### 3.1. Quantum model

With quantum-chemical descriptors, 18 descriptors were evicted in the reached model which are listed in table 1 while their numerical values are in equation $A_{calc}$ (9).

$$A_{calc} = -134.3 + 0.211D_{iq} - 12.73D_{2iq} + 175.5D_{3iq} + 1771D_{4iq} + 0.06162D_{5iq}$$

$$+ 0.28D_{20} - 0.151D_{21} + 0.191D_{22} + 1.55D_{23} + 0.712D_{24} - 0.2957D_{25}$$

$$- 0.438D_{26} - 0.171D_{27} + 0.111D_{28} - 0.152D_{29} + 4.55D_{30} + 1451D_{31}$$

$$- 0.157D_{32}$$

(9)

$R^2 = 0.9898$, $F$-statistic = 30.79, $R'^2$adj = 0.9577, $N = 26$.

### Table 4. A dataset of 26 N-arylantranilinic acids with calculated anti-inflammatory activity

| No. | R₁ | R₂ | R₃ | R₄ | R₅ | MEMP | $A_{exp}$ | $A_{calc}$ | $A_{exp} - A_{calc}$ |
|-----|----|----|----|----|----|------|-----------|------------|---------------------|
| 1   | Cl | H  | Cl | H  | Cl | 0.8  | 3.699    | 3.725     | -0.026             |
| 2   | CH₃ | SO,N(CH₃) | H | H  | CH₃ | 0.5  | 3.903    | 3.902     | 0.001              |
| 3   | CH₃ | NH₂ | H  | H  | Cl  | 6.2  | 2.809    | 2.863     | -0.054             |
| 4   | CH₃ | CH₃ | H  | H  | Cl  | 12.5 | 2.505    | 2.546     | -0.041             |
| 5   | Cl  | Cl  | H  | H  | CH₃ | 0.8  | 3.699    | 3.525     | 0.174              |
| 6   | Cl  | H  | CH₃ | H  | Cl  | 0.8  | 3.699    | 3.529     | 0.170              |
| 7   | Cl  | H  | Cl  | Cl | H  | 400  | 1.000    | 0.919     | 0.081              |
| 8   | Cl  | Cl  | Cl | H  | H  | 200  | 1.301    | 1.404     | -0.103             |
| 9   | Cl  | H  | Cl  | H  | Cl  | 100  | 1.602    | 1.747     | -0.145             |
| 10  | NH₂ | CH₃ | H  | H  | CH₃ | 2.5  | 2.204    | 2.180     | 0.024              |
| 11  | CH₃ | CH₃ | H  | H  | CH₃ | 6.2  | 2.809    | 2.755     | 0.054              |
| 12  | Cl  | CH₃ | H  | H  | CH₃ | 3.1  | 3.110    | 3.291     | -0.181             |
| 13  | CH₃ | Cl  | H  | CH₃ | H  | 1.6  | 3.397    | 3.369     | 0.028              |
| 14  | CH₃ | CH₃ | H  | H  | CH₃ | 1.6  | 3.397    | 3.351     | 0.046              |
| 15  | CH₃ | NH₂ | H  | H  | Cl  | 1.3  | 3.488    | 3.455     | 0.033              |
| 16  | CH₃ | SO,N(CH₃) | H | H  | CH₃ | 0.6  | 3.823    | 3.829     | -0.006             |
| 17  | Cl  | N(CH₃) | H | H  | Cl  | 0.6  | 3.823    | 3.822     | 0.001              |
| 18  | CH₃ | SO,N(CH₃) | H | H  | CH₃ | 0.5  | 3.903    | 4.078     | -0.175             |
| 19  | Cl  | Cl  | Cl | H  | CH₃ | 12.5 | 2.505    | 2.389     | 0.116              |
| 20  | CH₃ | CH₃ | H  | CH₃ | H  | 100  | 1.602    | 1.611     | -0.009             |
| 21  | Cl  | Cl  | Cl | H  | Cl  | 12.5 | 2.505    | 2.613     | -0.108             |
| 22  | Cl  | CH₃ | H  | Cl  | 12.5 | 2.505  | 2.505    | 0.000              |
| 23  | Cl  | Cl  | Cl | Cl | Cl  | 100  | 1.602    | 1.607     | -0.005             |
| 24  | Cl  | H  | Cl  | Cl | Cl  | 1.6  | 3.397    | 3.397     | 0.000              |
| 25  | Cl  | Cl  | Cl | Cl | Cl  | 25   | 2.204    | 2.091     | 0.113              |
| 26  | CH₃ | CH₃ | Cl | CH₃ | Cl  | 100  | 1.602    | 1.588     | 0.014              |
The dataset N-arylanthranilic acids (1-26), the model shows the best correlation with $R^2=0.989$ (Figure 2)

**Figure 2.** The plot $A_{\text{cal}}$ v.s $A_{\text{exp}}$

### 3.2. Structural Molecular Fragment (SMF) model

ISIDA generates 582 predefined fragments (SMF), but among these 582 fragments, 24 contributed to build our model. The contribution matrix $M_{ij}$

$$M_{ij} = 26 \times 24$$

| SMF | $a_i$ |
|-----|------|
| 1   | C-C=O | SMF$_1$ | -0.398275 |
| 2   | C-N-C | SMF$_2$ | 0.204297  |
| 3   | C-C-C-N | SMF$_3$ | 1.236656  |
| 4   | C=C-C-N | SMF$_4$ | -0.493310 |
| 5   | C-C-C-O | SMF$_5$ | 1.447280  |
| 6   | C-C=C-C-N | SMF$_6$ | -0.731878 |
| 7   | C=C-C-C-C | SMF$_7$ | -0.338618 |
| 8   | C-C=C-C-F | SMF$_8$ | -0.887086 |
| 9   | C=C-C-C-F | SMF$_9$ | 0.597106  |
| 10  | C-C=C-N-C-C-C-O | SMF$_{10}$ | 1.220001 |
| 11  | C=C-C-N-C=C-C-C-F | SMF$_{11}$ | 0.675218 |
| 12  | Cl-C-C-C | SMF$_{12}$ | -0.325778 |
| 13  | Cl-C=C-N-C-C-C-O | SMF$_{13}$ | 1.365737 |
| 14  | C-C-C-C | SMF$_{14}$ | 0.857632  |
| 15  | Cl-C=C-C-N-C-C-C-C-C | SMF$_{15}$ | -0.402841 |
| 16  | Cl-C=C-N-C-C-C-O | SMF$_{16}$ | 0.916635  |
| 17  | S-C-C-C | SMF$_{17}$ | 1.015195  |
| 18  | Cl-C-C-C | SMF$_{18}$ | 0.453626  |
| 19  | Cl-C-C-Cl | SMF$_{19}$ | -0.470649 |
| 20  | Cl-C-C-N-C | SMF$_{20}$ | 0.718064  |
| 21  | Cl-C=C-C-Cl | SMF$_{21}$ | 0.620912  |
| 22  | C-C-C-C-C-C-C-C | SMF$_{22}$ | -0.378501 |
| 23  | C=C-C=C-C-C-C | SMF$_{23}$ | 0.845447  |
| 24  | S-C=C-C-C-C-C | SMF$_{24}$ | 0.794212  |

From the contribution matrix and table (5), we will express the predicted activity as a linear function of Structural Molecular Fragment. Here we have 26 equations, so one for each molecule. Let express the linear equation of molecule number 1.

\[
A_{\text{act,1}} = a_{1b} + 0.398 \times SMF_1 + 0.204 \times SMF_2 + 1.236 \times SMF_3 - 0.493 \times SMF_4 + 1.447 \times SMF_5 - 0.731 \times SMF_6 - 0.338 \times SMF_7 - 0.887 \times SMF_8 + 0.597 \times SMF_9 - 0.325 \times SMF_{10} + 0.916 \times SMF_{11} + 0.343 \times SMF_{12} 
\]

\[
A_{\text{act,1}} = 3.690000
\]
The predicted activity of all these molecules is confined in table (6) below:

**Table 6.** A dataset of 26 N-arylanthranilic acids with calculated anti-inflammatory activity

| Mol | R1 | R2 | R3  | R4  | R5  | MED | AEXP | ACAL | AEXP - ACAL |
|-----|----|----|-----|-----|-----|-----|------|------|-------------|
| 1   | Cl | H  | CF3 | H   | Cl  | 0.8 | 3.699| 3.699000| 0.000000    |
| 2   | CH3| SO2N(CH3)2| H | H   | CH3| 0.5 | 3.903| 4.076463| -0.176463   |
| 3   | CH3| NH2| H   | H   | Cl  | 6.2 | 2.809| 2.638972| 0.161028    |
| 4   | CH3| CH3| H   | H   | Cl  | 12.5| 2.505| 2.902764| -0.402764   |
| 5   | Cl | Cl | H   | H   | CH3| 0.8 | 3.699| 3.561545| 0.128455    |
| 6   | Cl | H  | C2H5| H   | Cl  | 0.8 | 3.699| 3.512326| 0.177674    |
| 7   | Cl | H  | Cl  | Cl  | H   | 400 | 1.000| 1.190064| -0.190064   |
| 8   | Cl | Cl | Cl  | H   | Cl  | 200 | 1.301| 1.934889| -0.634889   |
| 9   | Cl | Cl | Cl  | H   | Cl  | 100 | 1.602| 1.757864| -0.157864   |
| 10  | NH2| CH3| H   | H   | CH3| 3.1 | 3.110| 3.216865| -0.106865   |
| 11  | Cl | CH3| H   | H   | Cl  | 3.1 | 3.110| 2.996924| 0.393076    |
| 12  | Cl | CH3| H   | H   | Cl  | 1.6 | 3.390000| 2.996924| 0.393076    |
| 13  | Cl | CH3| H   | CH3| H   | 1.6 | 3.397| 3.210410| 0.179590    |
| 14  | Cl | C2H5| H | H   | CH3| 1.6 | 3.397| 3.251863| 0.228137    |
| 15  | Cl | CH3| H   | H   | Cl  | 1.3 | 3.488| 3.216865| 0.228137    |
| 16  | Cl | Cl | Cl  | H   | Cl  | 1.3 | 3.488| 3.016044| 0.228137    |
| 17  | Cl | Cl | H   | C2H5| H   | 1.3 | 3.488| 3.016044| 0.228137    |
| 18  | Cl | Cl | Cl  | H   | Cl  | 1.3 | 3.488| 3.016044| 0.228137    |
| 19  | Cl | Cl | CH3| H   | Cl  | 1.3 | 3.488| 3.016044| 0.228137    |
| 20  | Cl | Cl | H   | Cl  | Cl  | 1.6 | 3.488| 3.016044| 0.228137    |
| 21  | Cl | Cl | Cl  | Cl  | H   | 1.6 | 3.488| 3.016044| 0.228137    |
| 22  | Cl | Cl | Cl  | Cl  | Cl  | 1.6 | 3.488| 3.016044| 0.228137    |
| 23  | Cl | Cl | Cl  | Cl  | Cl  | 25  | 2.608| 2.691396| 0.082404    |
| 24  | Cl | Cl | Cl  | Cl  | Cl  | 12.5| 2.505| 2.327175| 0.172825    |
| 25  | Cl | Cl | Cl  | Cl  | Cl  | 12.5| 2.505| 2.035218| 0.464782    |
| 26  | Cl | Cl | Cl  | Cl  | Cl  | 25  | 2.608| 2.151661| 0.048339    |

The quantum model give a squared correlation coefficient ($R^2=0.9898$), a Root Mean Square Error (RMSE=0.090), an adjusted R-squared ($R^2_{adj} = 0.9577$) and Mean Average Error (MAE= 0.065), also The ISIDA model give a squared correlation coefficient ($R^2=0.9077$), a Root Mean Square Error (RMSE= 0.277), an adjusted R-squared ($R^2_{adj} = 0.8351$) and Mean Average Error (MAE=0.206). See table (7)
Table 7. The results of QSAR analysis by MLR for Quantum model and ISIDA model for the compounds in series 1-26

|                      | MLR Equation | n | R²     | R²(adj) | MAE  | RMSE  |
|----------------------|--------------|---|--------|---------|------|-------|
| Quantum model        | Equation (9)  | 26| 0.9898 | 0.9577  | 0.065| 0.090 |
| ISIDA model          | Equation (10) | 26| 0.9077 | 0.8351  | 0.206| 0.277 |

4. CONCLUSION
The QSAR studies were conducted with a series of 26 structures of N-arylanthranilic acid and some useful predictive molecular models were obtained. The physicochemical descriptors were found to have an important role in the determining of the activity. To test the robustness of these models, we evaluate the coefficient of correlation, $R^2$, which defines the degree of dependence between theoretical and experimental variables. Two models were studied, quantum model and Structural Molecular Fragment (SMF) model. The two models present a good coefficient correlation, 0.989 and 0.907 respectively. Among the two QSAR models (quantum and ISIDA), results of quantum analysis showed significant predictive power. But the advantages of Structural Molecular Fragment model are the power of fragment descriptors originates from their universality, very high computational efficacy, simplicity of interpretation, as well as their high diversity and versatility.

ACKNOWLEDGEMENTS
Part of this work was supported by international foundation of science (No. F/4893-1) and the third world academy science (No.10.004RG/AC-I). SFK also thank the Humboldt Foundation for equipment.

REFERENCES
[1] J.H. Botting. Nonsteroidal antiinflammatory agents. Drugs today (1999), 35, 225-235. View Article
[2] S. Prasanna, E. Manivannan, S.C. Chaturvedi, QSAR analyses of conformationally restricted 1,5-diaryl pyrazoles as selective COX-2 inhibitors: application of connection table representation of ligands, Bioorg & Med Chem Let, (2005), 15, 8, 2097-2102. View Article
[3] J.R. Vane and R.M. Botting, new insights into the mode of action of anti-inflammatory drugs. Inflamm. Res. 44(1995) 1-10. View Article
[4] J. G. Hardman and L. E. Limbird (Eds.), Goodman & Gilman's The Pharmacological Basis of Therapeutics, Int. Ed McGraw-Hill, USA, (1996), 635.
[5] Edmond C. Ku, Warren Lee, Himanshu V. Kothari, Dietrich W. Scholer, Effect of diclofenac sodium on the arachidonic acid cascade. The American Journal of Medicine. Volume 80, Issue 4, 28 April (1986), Pages 18-29074-4 View Article
[6] C. Almansa, J. Alfon, A. F. Arriba, F. L. Cavalcanti, I. Escamilla, L. A. Gomez, A. Miralles, R. Solvia, J. Bartoli, E. Caracelli, M. Merlos, and J. G. Rafanell, Synthesis and Structure-Activity Relationship of a New Series of COX-2 Selective Inhibitors: 1,5-Diarylimidazoles. Med. Chem.46 (2003) 3463-3475. View Article
[7] J. R. Mclean and M. I. Gluckman, Arzneim. On the mechanism of the pharmacologic activity of meclofenamate sodium. Forsch.33 (1983) 627-631
[8] H. Kankaanranta, M. Luomala, A. Kosonen, and E. Moilanen, Br. Inhibition by fenamates of calcium influx and proliferation of human lymphocytes J. Pharmacol.119(1996) 487-494. View Article
[9] B.K. Tiwari, A. Thakur, M. Thakur, N. D. Pande, S. S. Narvi, and S. Thakur. Modeling of cytotoxicity on some non-nucleoside reverse transcriptase inhibitors of HIV-1: role of physicochemical parameters. ARKIVOC (2006) (ii) 213-225 View Article
[10] Vesna Dimova, Katka Colanceska Ragenovic and Vladimir Kakurinov. QSAR of Some N-Aryl/Heteroarylaminomethyl/ethyl-1,2,4-Triazoles Part II: Antimicrobial Activity Against Bacillus Subtilis. Int. J. Mol. Sci. (2006), 7, 119-129. View Article
[11] Sofie Van Damme. Quantum chemistry in QSAR Quantum chemical descriptors Use, benefits and drawbacks. Thesis submitted in fulfillment of the requirements for the degree of Doctor (Ph.D.) in Sciences: Chemistry, 2009
[12] Ojha Lokendra Kumar, Chaturvedi Ajay M, Bhardwaj Arpan. QSAR study of some TIBO derivatives: A non conventional topological pa-
rameter approach. Int. J. Res. Chem. Env. (2011), 16-23

[13]A. Beheshti, P. Norouzi, M. R. Ganjali. A Simple and Robust Model for Predicting the Reduction Potential of Quinones Family; Electrophilicity Index Effect. Int. J. Electrochem. Sci.,7(2012), 4811-4821.

[14]Gustavo P. ROMANELLI, Juan Carlos AUTINO, Eduardo A. CASTRO. Application of Quantitative Structure-Retention Relationships (QSRR) to a Set of Organic Bromo and Nitrile Derivatives. Turk J Chem 26 (2002), 335-343.

[15]M. Karelson; V.S. Lobanov; A.R. Katritzky, Quantum-Chemical. Descriptors in QSAR/QSPR Studies. Chem Rev, (1996), 96, 1027-1043. View Article

[16]Ana M. Pajor; Nina N. Sun. nonsteroidal anti-inflammatory drugs and other anthranilic acids inhibit the Na+/ dicarboxylate symporter from taphylococcus aureus. Biochemistry (2013), 52, 17, 2924-2932. View Article

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

[18]www.chemaxon.com

[19]Solov'ev V., Sukhno I., Buzko V., Polushin A., Marcou G., Tsivadze A., Varnek A. Stability Constants of Complexes of Zn2+, Cd2+, and Hg2+ with Organic Ligands : QSPR Consensus Modeling and Design of New Metal Binders. J. Incl. Phenom. Macrocycl. Chem., 2011, View Article

[20]Solov'ev V., Oprisiu I., Marcou G., Varnek A. Quantitative Structure - Property Relationship (QSPR) Modeling of Normal Boiling Point Temperature and Composition of Binary Azeotropes. Ind. Eng. Chem. Res., 2011, 50, No. 24, pp 14162-14167. View Article

[21]Varnek A., Solov'ev V. Quantitative Structure-Property Relationships in solvent extraction and complexation of metals. Rev. in Book : Ion Exchange and Solvent Extraction, A Series of Advances. Vol. 19, P. 319-358. A. K. Sengupta and B. A. Moyer, Eds., CRC Press, Taylor and Francis Group : Boca Raton, 2009, 679 pp. View Article

[22]Solov'ev, V. P. ; Varnek, A. A. ; Wipff, G. Modelling of Ion Complexation and Extraction Using Substructural Molecular Fragments. J. Chem. Inf. Comput. Sci., 2000, 40, P. 847-858. View Article

[23]Varnek, A. A. ; Wipff, G. ; Solov'ev, V. P. Towards an Information System on Solvent Extraction. J. Solvent Extr. Ion. Exch., 2001, 19, No. 5, P.791-837. View Article

[24]Varnek, A. A. ; Wipff, G. ; Solov'ev, V. P., Solotnov A.F. Assessment of The Macrocyclic Effect for The Complexation of Crown-Ethers with Alkali Cations Using the Substructural Molecular Fragments Method. J. Chem. Inf. Comput. Sci., 2002, 42, No. 4, P. 812-829. View Article

[25]Solov'ev, V. P. ; Varnek, A. Anti-HIV Activity of HEPT, TIBO and Cyclic Urea Derivatives : Structure-Property Studies, Focused Combinatorial Library Generation and Hits Selection Using Substructural Molecular Fragments Method. J. Chem. Inf. Comp. Sci., 2003, 43, No. 5, P. 1703-1719. View Article

[26]Katritzky, A.R. ; Fara, D.C. ; Yang, H. ; Karelson, M. ; Suzuki, T. ; Solov'ev, V. P. ; Varnek A. Quantitative Structure-Property Relationship Modeling of ?-Cycloextrin Complexation Free Energies. J. Chem. Inf. Comput. Sci. 2004, 44, No. 2, 529-541. View Article

[27]Varnek, A. ; Fourches, D. ; Solov'ev, V. P. ; Bauvin, V. E. ; Turanov, A. N. ; Karandashev, V. K. ; Fara, D. ; Katritzky, A. R. « In Silico » Design of New Uranyl Extractants Based on Phosphoryl-Containing Podands : QSPR Studies, Generation and Screening of Virtual Combinatorial Library and Experimental Tests. J. Chem. Inf. Comput. Sci., 2004, 44, No. 4, 1365-1382. View Article

[28]Varnek, A. ; Solov'ev, V. P. « In Silico » Design of Potential Anti-HIV Actives Using Fragment Descriptors. Combinatorial Chem. High Throughput Screening, 2005, 8, No. 5, 403-416. View Article

[29]Katritzky, A. R. ; Kumar, M. ; Fara, D. C. ; Karelson, M. ; Acree, W. E. Jr. ; Solov'ev, V. P. ; Varnek A. QSAR modeling of blood:air and tissue:air partition coefficients using theoretical descriptors. Bioorg. Med. Chem.,2005, 13, 6450-6463. View Article
[30] Katritzky, A. R.; Dobchev, D. A.; Fara, D. C.; Hur, E.; Tamrn, K.; Kuruneci, L.; Karelson, M.; Varnek, A.; Solov'ev, V. P. Skin Permeation Rate as a Function of Chemical Structure. J. Med. Chem., 2006, 49, No. 11, 3305-3314. View Article

[31] Katritzky, A. R.; Kuanar, M.; Slavov, S.; Dobchev, D. A.; Fara, D. C.; Karelson, M.; William, E.; Acree, W. E. Jr.; Solov'ev, V. P.; Varnek, A. Correlation of Blood - Brain Penetration Using Structural Descriptors. Bioorg. Med. Chem., 2006, 14, No. 14, 4888-4917. View Article

[32] Varnek, A.; Fourches, D.; Sieffert, N.; Solov'ev, V. P.; Hill, C.; Lecomte, M. QSAR Modeling of the AmIII / EuIII Separation Factor : How Far Can We Predict ? Solv. Extr. Ion Exch., 2007, 25, No. 1, P. 1-26. View Article

[33] Horvath D., Bonachera F., Solov'ev V., Gaudin C., Varnek A. Stochastic versus Stepwise Strategies for Quantitative Structure- Activity Relationship Generations. - How Much Effort May the Mining for Successful QSAR Models Take? J. Chem. Inf. Model., 2007, 47, No. 3, P. 927-939. View Article

[34] Varnek A.; Fourches D.; Solov'ev V.; Klimchuk O.; Ouadi A.; Billard I. Successful « In Silico » Design of New Efficient Urankyl Binders. Solv. Extr. Ion Exch., 2007, 25, No. 4, P. 433-462. View Article

[35] Varnek A.; Fourches D.; Horvath D.; Klimchuk O., Gaudin C., Vayer P., Solov'ev V., Hoonakker F., Tetko I. V., Marcou G. ISIDA - Platform for Virtual Screening Based on Fragment and Pharmacophoric Descriptors. Cur. Computer-Aided Drug Design, 2008, 4, No. 3, P. 191-198. View Article

[36] Varnek A., Fourches D., Kireeva N., Klimchuk O., Marcou G., Tsivadze A., Solov'ev V. Computer-Aided Design of New Metal Binders. Radiochim. Acta, 2008, 96, P. 505-511. View Article

[37] A Varnek , D Fourches, F Hoonakker, V P Solov'ev. Substructural fragment: an universal language to encode reactions, molecular and supramolecular structures. J. Comput. Aided Mol. Des., (2005), 19(9-10): P.693-703. View Article

[38] U.Oleg, Diudea, M.V. activity prediction by cluj-simil program. Rev. Roum. Chim. (2003),48 (4):321-330

[40] Solov'ev, V.; Varnek, A.; Tsivadze, A., QSAR Ensemble Modelling of the 1:1 and 1:2 Complexation of Co2+, Ni2+, and Cu2+ with Organic Ligands. Relationships between Stability Constants. J. Comput. Aided Mol. Des. (2014), 28, (5), 549-564. View Article

[41] Varnek, A.; Kireeva, N.; Tetko, I. V.; Baskin, I. I.; Solov'ev, V. P., Exhaustive QSAR Studies of a Large Diverse Set of Ionic Liquids: How Accurately Can We Predict Melting Points? J. Chem. Inf. Model. (2007), 47, (3), 1111-1122. View Article

[42] Tetko, I. V.; Solov'ev, V. P.; Antonov, A. V.; Yao, X. J.; Fan, B. T.; Hoonakker, F.; Fourches, D.; Lachiche, N.; Varnek, A., Benchmarking of Linear and Non-Linear Approaches for Quantitative Structure-Property Relationship Studies of Metal Complexation with Organic Ligands. J. Chem. Inf. Model. (2006), 46, (2), 808-819. View Article