Quantitative structure activity relationship of aroyl-pyrrolyl-hydroxy-amide derivatives

Tajuddin Khan*, Suhail Ahmad Khan
Department of Chemistry, M. L. K. (P. G.) College, Balrampur, U. P., (INDIA)
E-mail: tkhan1250@yahoo.com
Received: 10th April, 2012 ; Accepted: 10th May, 2012

ABSTRACT
In this article we have done QSAR study of ten derivative of Aroyl-Pyrrolyl-Hydroxy-Amide with the help of Quantum chemical parameters. For QSAR prediction, the 3D modeling and geometry optimization of all the compounds have been done with the help of PCModel software using PM3 hamiltonian. The MOPAC calculations have been performed with WINMOPAC 7.21 software, by applying keywords PM3 Charge=0 Gnorm=0.1, Bonds, Geo-OK, Vectors density. Various QSAR model have been obtained with the help of MLR analysis. Only ten models are reported on the basis of correlation coefficient. Model no. ten has been recognized is the best model on the basis of statistical parameters. This model has been evaluated by molecular weight, total energy, HOMO energy and absolute hardness.

INTRODUCTION
QSAR study of phenols with the help of quantum mechanical parameter has recently been made by Singh et al[1]. They developed QSAR models having high degree of predicted power with correlation coefficient value above 0.88. QSAR[2-5] has become increasingly helpful in understanding many aspect of chemical biological activity in drug research and pharmacological sciences[6]. In this present work we have done quantitative structure activity relationship study of some novel heterocyclic compound i.e. Aroyl-, Pyrrolyl-, Hydroxy-Amide derivatives with the help of quantum chemical descriptors. Various QSAR model are obtained by multiple linear regression analysis. Only ten QSAR model have been selected on the basis of correlation coefficient.

EXPERIMENTAL
The study material for this paper are ten Aroyl-Pyrrolyl-, Hydroxy-Amide derivatives which have been arranged on the basis of 50% inhibitory concentration (IC\textsubscript{50}). For QSAR study we have been used following descriptors:
1- Molecular weight (M\textsubscript{W})
2- Heat of formation (H\textsubscript{f0})
3- Total energy (E\textsubscript{T})
4- HOMO value (\varepsilon\textsubscript{HOMO})
5- LUMO value (\varepsilon\textsubscript{LUMO})
6- Electronegativity (\chi)
7- Absolute hardness (\eta)
For QSAR prediction, the 3D modeling and geometry optimization of all the compounds have been done with the help of PCModel software using PM3 hamiltonian[7]. The MOPAC calculations have been performed with WINMOPAC 7.21 software, by applying keywords PM3 Charge=0 Gnorm=0.1, Bonds, Geo-OK, Vectors density. The values of the above descriptors are calculated by solving the equations given below:
In DFT, the electronegativity, commonly known to a chemist, is defined as the negative of a partial derivative of energy $E$ of an atomic or molecular system with respect to the number of electrons $N$ with a constant external potential $v(r)^8$

$$\mu = \chi = - (\delta E / \delta N)_{v(r)}$$  \hspace{1cm} (1)

In accordance with the earlier work of Iczkowski and Margrave$^9$, it should be stated that when assuming a quadratic relationship between $E$ and $N$ and in a finite difference approximation, Eq. 1 may be rewritten as

$$\chi = -\mu = -(IE + EA)/2$$  \hspace{1cm} (2)

where $IE$ and $EA$ are the vertical ionization energy and electron affinity, respectively, thereby recovering the electronegativity definition of Mulliken$^{10}$. Moreover, a theoretical justification was provided for Sanderson's principle of electronegativity equalization, which states that when two or more atoms come together to form a molecule, their electronegativities become adjusted to the same intermediate value$^{11-13}$.

The absolute hardness $\eta$ is defined as$^{14}$

$$\eta = (IP- EA)/2$$  \hspace{1cm} (3)

where $IP$ and $EA$ are the ionization potential and electron affinity respectively, of the chemical species. According to the Koopman's theorem, the $IP$ is simply the energy of the HOMO with change of sign and the $EA$ is the eigenvalue of the LUMO with change of sign hence the equations 2 and 3 can be written as

$$\chi = (\varepsilon_{HOMO} + \varepsilon_{LUMO})/2$$  \hspace{1cm} (4)

$$\eta = (\varepsilon_{HOMO} - \varepsilon_{LUMO})/2$$  \hspace{1cm} (5)

The heat of formation is defined as:

$$\Delta H_f^o = E_{elect} + E_{nuc} - E_{isol} + E_{atom}$$  \hspace{1cm} (6)

where $E_{elect}$ is the electronic energy, $E_{nuc}$ is the nuclear-nuclear repulsion energy, $E_{isol}$ is the energy required to strip all the valence electrons of all the atoms in the system, and $E_{atom}$ is the total heat of atomization of all the atoms in the system. The total electronic energy of the system is given by$^{15}$

$$E_e = 1/2 \mathbf{R}(\mathbf{H}+\mathbf{F})$$  \hspace{1cm} (7)

where $\mathbf{R}$ is the density matrix and $\mathbf{H}$ is the one-electron matrix. $\mathbf{F}$ is fock matrix.

For regression analysis, we used project leader programme associated with CAChe pro software of Fujitsu. Various regression equation were developed for the prediction of activity. The values of descriptors have been calculated by the above equation.

**RESULT AND DISCUSSION**

The parent skeleton of Aroyl-Pyrroyl-Hydroxy-Amide$^{16}$ is given below:

![Parent skeleton of Aroyl-Pyrroyl-Hydroxy-Amide](image)

The 10 derivatives of Aroyl-Pyrroyl-Hydroxy-Amide and their observed biological activity in terms of inhibitory growth concentration (PIC$^{50}$). The values of various quantum chemical descriptors of Aroyl-Pyrroyl-Hydroxy-Amide derivatives, along their observed biological activity are placed in **TABLE 1**. The quantities of descriptors in a number of combinations have been used for MLR analysis and QSAR models. Out of them only

**TABLE 1 : Aroyl-pyrroyl-hydroxy-amide derivatives with their observed activity**

<table>
<thead>
<tr>
<th>Comp.No.</th>
<th>$R$</th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>IC$^{50}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CH$_3$</td>
<td>Ph</td>
<td></td>
<td>3.8</td>
</tr>
<tr>
<td>2</td>
<td>CH$_3$</td>
<td>4-Cl-Ph</td>
<td></td>
<td>2.4</td>
</tr>
<tr>
<td>3</td>
<td>CH$_3$</td>
<td>4-F-Ph</td>
<td></td>
<td>3.8</td>
</tr>
<tr>
<td>4</td>
<td>CH$_3$</td>
<td>4-O$_2$N-Ph</td>
<td></td>
<td>3.9</td>
</tr>
<tr>
<td>5</td>
<td>4CH$_3$.Ph</td>
<td></td>
<td></td>
<td>1.9</td>
</tr>
<tr>
<td>6</td>
<td>CH$_3$</td>
<td>4-CH$_2$.OPh</td>
<td></td>
<td>2.9</td>
</tr>
<tr>
<td>7</td>
<td>CH$_3$</td>
<td>4Me$_2$.N-Ph</td>
<td></td>
<td>2.4</td>
</tr>
<tr>
<td>8</td>
<td>CH$_3$</td>
<td>Ph-CH$_3$</td>
<td></td>
<td>0.1</td>
</tr>
<tr>
<td>9</td>
<td>CH$_3$</td>
<td>Ph.CH === CH</td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>10</td>
<td>H</td>
<td>Ph</td>
<td></td>
<td>5.0</td>
</tr>
</tbody>
</table>
10 QSAR models presented below, have been found to have very high-predicted power. The predicted activities of these QSAR models are placed in TABLE 3.

**QSAR model**

The predicted activity PA₁ is calculated by solving regression equation 1

\[
PA₁ = -0.000783155 * H_f + 0.00515297 * MW - 1.89565 \\
\text{rCV}^2 = 0.796811 \\
\text{r}^2 = 0.813363
\]

Equation 1 involves heat of formation as first descriptor and molecular weight as second descriptor. The values of quantum chemical descriptors are presented in TABLE 2. The reliability of this regression model can be tested from correlation coefficient 0.813363 and cross validation coefficient 0.796811. These coefficients indicate that this regression gives good regression result. The values of the predicted activities PA₁ are listed in the TABLE 3.

**TABLE 2 : Values of various quantum chemical descriptors with observed activity (IC_{50})**

<table>
<thead>
<tr>
<th>No.</th>
<th>MW</th>
<th>H_f</th>
<th>E_T</th>
<th>(\varepsilon)HOMO</th>
<th>(\varepsilon)LUMO</th>
<th>(\chi)</th>
<th>(\eta)</th>
<th>IC_{50}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>270.287</td>
<td>-42.788</td>
<td>-146.753</td>
<td>-9.101</td>
<td>-0.774</td>
<td>-4.937</td>
<td>4.163</td>
<td>3.8</td>
</tr>
<tr>
<td>2</td>
<td>256.26</td>
<td>-44.933</td>
<td>-139.63</td>
<td>-9.173</td>
<td>-0.793</td>
<td>-4.983</td>
<td>4.19</td>
<td>2.4</td>
</tr>
<tr>
<td>3</td>
<td>304.732</td>
<td>-49.462</td>
<td>-143.74</td>
<td>-9.156</td>
<td>-0.818</td>
<td>-4.987</td>
<td>4.169</td>
<td>3.8</td>
</tr>
<tr>
<td>4</td>
<td>288.278</td>
<td>-87.011</td>
<td>-158.23</td>
<td>-9.187</td>
<td>-0.841</td>
<td>-5.014</td>
<td>4.173</td>
<td>3.9</td>
</tr>
<tr>
<td>5</td>
<td>315.285</td>
<td>-70.266</td>
<td>-178.569</td>
<td>-9.405</td>
<td>-1.349</td>
<td>-5.377</td>
<td>4.028</td>
<td>1.9</td>
</tr>
<tr>
<td>6</td>
<td>284.314</td>
<td>-52.425</td>
<td>-153.941</td>
<td>-9.081</td>
<td>-0.76</td>
<td>-4.921</td>
<td>4.16</td>
<td>2.9</td>
</tr>
<tr>
<td>7</td>
<td>300.313</td>
<td>-83.312</td>
<td>-166.13</td>
<td>-9.081</td>
<td>-0.763</td>
<td>-4.922</td>
<td>4.159</td>
<td>2.4</td>
</tr>
<tr>
<td>8</td>
<td>313.355</td>
<td>-49.261</td>
<td>-170.447</td>
<td>-8.649</td>
<td>-0.725</td>
<td>-4.687</td>
<td>3.962</td>
<td>0.1</td>
</tr>
<tr>
<td>9</td>
<td>284.314</td>
<td>-52.186</td>
<td>-153.941</td>
<td>-9.086</td>
<td>-0.764</td>
<td>-4.925</td>
<td>4.161</td>
<td>1</td>
</tr>
</tbody>
</table>

**TABLE 3 : Values of predicted activities from PA1 to PA10**

<table>
<thead>
<tr>
<th>Compd.No.</th>
<th>PA₁</th>
<th>PA₂</th>
<th>PA₃</th>
<th>PA₄</th>
<th>PA₅</th>
<th>PA₆</th>
<th>PA₇</th>
<th>PA₈</th>
<th>PA₉</th>
<th>PA₁₀</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.639</td>
<td>2.338</td>
<td>2.546</td>
<td>2.711</td>
<td>2.628</td>
<td>2.693</td>
<td>2.498</td>
<td>2.502</td>
<td>2.557</td>
<td>2.633</td>
</tr>
<tr>
<td>2</td>
<td>2.492</td>
<td>2.163</td>
<td>2.403</td>
<td>2.72</td>
<td>2.468</td>
<td>2.63</td>
<td>2.356</td>
<td>2.547</td>
<td>2.584</td>
<td>2.564</td>
</tr>
<tr>
<td>3</td>
<td>2.831</td>
<td>2.971</td>
<td>2.641</td>
<td>3.227</td>
<td>2.837</td>
<td>2.919</td>
<td>3.993</td>
<td>3.328</td>
<td>2.737</td>
<td>2.616</td>
</tr>
<tr>
<td>4</td>
<td>2.165</td>
<td>2.561</td>
<td>2.737</td>
<td>2.076</td>
<td>2.193</td>
<td>2.053</td>
<td>2.756</td>
<td>3.689</td>
<td>2.616</td>
<td>3.328</td>
</tr>
<tr>
<td>5</td>
<td>2.625</td>
<td>1.299</td>
<td>1.608</td>
<td>3.731</td>
<td>2.493</td>
<td>3.345</td>
<td>1.505</td>
<td>2.114</td>
<td>2.055</td>
<td>1.974</td>
</tr>
<tr>
<td>6</td>
<td>2.622</td>
<td>2.501</td>
<td>2.36</td>
<td>2.576</td>
<td>2.633</td>
<td>2.565</td>
<td>2.903</td>
<td>2.739</td>
<td>2.527</td>
<td>2.616</td>
</tr>
<tr>
<td>7</td>
<td>2.317</td>
<td>1.975</td>
<td>1.4</td>
<td>1.957</td>
<td>2.374</td>
<td>2.04</td>
<td>3.075</td>
<td>1.652</td>
<td>0.947</td>
<td>1.996</td>
</tr>
<tr>
<td>8</td>
<td>2.906</td>
<td>2.549</td>
<td>2.533</td>
<td>1.473</td>
<td>2.946</td>
<td>2.388</td>
<td>0.537</td>
<td>0.363</td>
<td>1.762</td>
<td>0.133</td>
</tr>
<tr>
<td>9</td>
<td>2.623</td>
<td>2.501</td>
<td>2.367</td>
<td>2.601</td>
<td>2.635</td>
<td>2.579</td>
<td>2.92</td>
<td>2.769</td>
<td>2.546</td>
<td>2.647</td>
</tr>
</tbody>
</table>

**QSAR model**

The predicted activity PA₂ is calculated by solving regression equation 2

\[
PA₂ = 0.00515297 * MW - 0.491437 * E_T - 1.89565 \\
\text{rCV}^2 = 0.796811 \\
\text{r}^2 = 0.813363
\]

Equation 2 involves molecular weight as first descriptor and total energy as second descriptor. The values of quantum chemical descriptors are presented in TABLE 2. The reliability of this regression model can be tested from correlation coefficient 0.813363 and cross validation coefficient 0.796811. These coefficients indicate that this regression gives good regression result. The values of the predicted activities PA₂ are listed in the TABLE 3.

**QSAR model**

The predicted activity PA₃ is calculated by solving regression equation 3

\[
PA₃ = -0.000488281 * H_f + 0.00515297 * MW - 0.0625 * E_T
\]

Equation 3 involves heat of formation as first descriptor and molecular weight as second descriptor. The values of quantum chemical descriptors are presented in TABLE 2. The reliability of this regression model can be tested from correlation coefficient 0.813363 and cross validation coefficient 0.796811. These coefficients indicate that this regression gives good regression result. The values of the predicted activities PA₃ are listed in the TABLE 3.
Equation 3 involves heat of formation as first descriptor and molecular weight as second descriptor. The third descriptor is total energy. The values of quantum chemical descriptors are presented in Table 2. The reliability of this regression model can be tested from correlation coefficient 0.811212 and cross validation coefficient 0.790595. These coefficients indicate that this regression gives good regression result. The values of the predicted activities PA_4 are listed in the Table 3.

**QSAR model**

The predicted activity PA_4 is calculated by solving regression equation 4

\[ PA_4 = -0.00071984* H_f + 0.00521657* MW - 0.0480385* eHOMO - 2.33504 \]

\[ r^2 = 0.811212 \]

Equation 4 involves heat of formation as first descriptor and the second and third descriptors is molecular weight and HOMO energy respectively. The values of quantum chemical descriptors are presented in Table 2. The reliability of this regression model can be tested from correlation coefficient 0.815118 and cross validation coefficient 0.764644. These coefficients indicate that this regression gives good regression result. The values of the predicted activities PA_4 are listed in the Table 3.

**QSAR model**

The predicted activity PA_5 is calculated by solving regression equation 5

\[ PA_5 = -0.000842964* H_f + 0.00534636* MW + 0.0670735* eLUMO - 1.9109 \]

\[ r^2 = 0.822225 \]

Equation 5 involves heat of formation as first descriptor and molecular weight as second descriptor. The third descriptor is LUMO energy. The values of quantum chemical descriptors are presented in Table 2. The reliability of this regression model can be tested from correlation coefficient 0.822225 and cross validation coefficient 0.744483. These coefficients indicate that this regression gives good regression result. The values of the predicted activities PA_5 are listed in the Table 4. Graph between the observed activity and the predicted activities PA_5 are listed in the Table 3.

**QSAR model**

The predicted activity PA_6 is calculated by solving regression equation 6

\[ PA_6 = -0.000765366* H_f + 0.00532853* MW + 0.0834568* e\chi - 2.28681 \]

\[ r^2 = 0.8204 \]

Equation 6 involves heat of formation as first descriptor and molecular weight as second descriptor. The third descriptor is electronegativity. The values of quantum chemical descriptors are presented in Table 2. The reliability of this regression model can be tested from correlation coefficient 0.8204 and cross validation coefficient 0.743636. These coefficients indicate that this regression gives good regression result. The values of the predicted activities PA_6 are listed in the Table 4.

### Table 4: Correlation summary of best QSAR models

<table>
<thead>
<tr>
<th>PA</th>
<th>rcV²</th>
<th>r²</th>
<th>VU</th>
<th>SE</th>
<th>SEE</th>
<th>t-Value</th>
<th>p-Value</th>
<th>DOF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.796811</td>
<td>0.813363</td>
<td>H_f, MW</td>
<td>0.996</td>
<td>1.477</td>
<td>1.0029</td>
<td>0.1726</td>
<td>0.0007</td>
</tr>
<tr>
<td>2</td>
<td>0.79681</td>
<td>0.813363</td>
<td>MW, E_T</td>
<td>0.594</td>
<td>1.313</td>
<td>1.843</td>
<td>0.0512</td>
<td>0.2105</td>
</tr>
<tr>
<td>3</td>
<td>0.790595</td>
<td>0.811212</td>
<td>H_f, MW, E_T</td>
<td>0.561</td>
<td>1.292</td>
<td>1.940</td>
<td>0.0442</td>
<td>0.2350</td>
</tr>
<tr>
<td>4</td>
<td>0.764644</td>
<td>0.815118</td>
<td>H_f, MW, eHOMO</td>
<td>0.543</td>
<td>1.313</td>
<td>1.841</td>
<td>0.0514</td>
<td>0.2098</td>
</tr>
<tr>
<td>5</td>
<td>0.744833</td>
<td>0.822225</td>
<td>H_f, MW, eLUMO</td>
<td>0.990</td>
<td>1.476</td>
<td>1.010</td>
<td>0.1709</td>
<td>0.0024</td>
</tr>
<tr>
<td>6</td>
<td>0.743636</td>
<td>0.8204</td>
<td>H_f, MW, e\chi</td>
<td>0.815</td>
<td>1.438</td>
<td>1.225</td>
<td>0.1276</td>
<td>0.0329</td>
</tr>
<tr>
<td>7</td>
<td>0.778457</td>
<td>0.819805</td>
<td>H_f, MW, \eta</td>
<td>0.282</td>
<td>0.977</td>
<td>3.543</td>
<td>0.0038</td>
<td>0.5629</td>
</tr>
<tr>
<td>8</td>
<td>0.764744</td>
<td>0.815118</td>
<td>MW, E_T, eHOMO</td>
<td>0.237</td>
<td>0.849</td>
<td>4.388</td>
<td>0.0012</td>
<td>0.6698</td>
</tr>
<tr>
<td>9</td>
<td>0.744483</td>
<td>0.822225</td>
<td>MW, E_T, \eta</td>
<td>0.441</td>
<td>1.195</td>
<td>2.398</td>
<td>0.0217</td>
<td>0.3455</td>
</tr>
<tr>
<td>10</td>
<td>0.730729</td>
<td>0.822328</td>
<td>MW, E_T, eHOMO, \eta</td>
<td>0.237</td>
<td>0.848</td>
<td>4.3966</td>
<td>0.0011</td>
<td>0.6707</td>
</tr>
</tbody>
</table>
The predicted activity $PA_7$ is calculated by solving regression equation 7

$$PA_7 = -0.000932053 \times H_f + 0.00525805 \times MW + 0.134769 \times \eta^{-1.29462}$$

$$r^2 = 0.778457$$

Equation 7 involves heat of formation as first descriptor and molecular weight as second descriptor. The third descriptor is absolute hardness. The values of quantum chemical descriptors are presented in TABLE 2. The reliability of this regression model can be tested from correlation coefficient 0.819805 and cross validation coefficient 0.778457. These coefficients indicate that this regression gives good regression result. The values of the predicted activities $PA_7$ are listed in the TABLE 3.

The predicted activity $PA_8$ is calculated by solving regression equation 8

$$PA_8 = 0.00521657 \times MW - 0.451707 \times E_T - 0.0480385 \times \epsilon\text{HOMO} - 2.33504 \times CV^2$$

$$r^2 = 0.764644$$

Equation 8 involves molecular weight as first descriptor and total energy as second descriptor. The third descriptor is HOMO energy. The values of quantum chemical descriptors are presented in TABLE 2. The reliability of this regression model can be tested from correlation coefficient 0.815118 and cross validation coefficient 0.764644. These coefficients indicate that this regression gives good regression result. The values of the predicted activities $PA_8$ are listed in the TABLE 3.

The predicted activity $PA_9$ is calculated by solving regression equation 9

$$PA_9 = 0.00534636 \times MW - 0.528968 \times E_T + 0.0670735 \times \eta^{-1.9109}$$

$$r^2 = 0.744483$$

Equation 9 involves molecular weight as first descriptor and total energy as second descriptor. The third descriptor is absolute hardness. The values of quantum chemical descriptors are presented in TABLE 2. The reliability of this regression model can be tested from correlation coefficient 0.822225 and cross validation coefficient 0.744483. These coefficients indicate that this regression gives good regression result. The values of the predicted activities $PA_9$ are listed in the TABLE 3.

The predicted activity $PA_{10}$ is calculated by solving regression equation 10

$$PA_{10} = 0.00534135 \times MW - 0.542908 \times E_T - 0.0578958 \times \epsilon\text{HOMO} + 0.143306 \times \eta^{-1.78611}$$

$$r^2 = 0.730729$$

Equation 10 involves molecular weight as first descriptor and total energy as second descriptor. The third and fourth descriptor are HOMO energy and absolute hardness. The values of quantum chemical descriptors are presented in TABLE 2. The reliability of this regression model can be tested from correlation coefficient 0.822328 and cross validation coefficient 0.730729. These coefficients indicate that this regression gives good regression result. The values of the predicted activities $PA_{10}$ are listed in the TABLE 3.

In this present study QSAR analysis of Aroyl-Pyrroyl-Hydroxy-Amide derivatives with the help of quantum chemical parameter. The best ten model have been selected. Out of ten QSAR model, model no. 10 is the best model on the basis of correlation coefficient which is greater than 0.822. The predicted activity is close to observed activity. This model is evaluated with the help of molecular weight, total energy, HOMO energy and absolute hardness, which is presented in TABLE 4.

**REFERENCES**