QSAR modeling of HIV-1 reverse transcriptase inhibitor of aryluracil derivatives using ab initio and empirical calculations

Neetu Sharma, Amrita Dwivedi, A K Srivastava & Ajeet Singh*
Department of Chemistry, University of Allahabad, Allahabad 211 002, India
E-mail: ajeetmolecule@gmail.com

Received 11 August 2015; accepted (revised) 8 February 2016

Quantitative structure–activity relationship (QSAR) studies have been performed on a series of 1-[(2-benzyloxyl/alkoxyl)methyl]-5-halo-6-aryluracil derivatives that are non-nucleoside reverse transcriptase inhibitors. Various descriptors have been calculated based on empirical and density functional theory (DFT). Density functional theory based descriptors have been calculated at GGA-PW91 level. Several QSAR equations have been formulated through regression analysis and tested with external and internal validation tests. The best equations have been selected from the various statistically significant equations. Model equations have been cross validated by leave one out (LOO) technique. The calculated results suggest that the introduction of a halogen at the R2 position may contribute to the effectiveness of these compounds against RTI-resistant variants.

Keywords: DFT, QSAR, regression analysis, molecular descriptor

The most communicable and fatal infection is human immunodeficiency virus (HIV) and almost every country is paying attention to preventing it. In scientific community, it is one of the unattainable targets because the virus has magic capacity and can mutate the gene sequence. Till date no remedial chemotherapy has been developed for HIV and it is commonly known as acquired immune deficiency syndrome (AIDS). This infection targets cells of immune system expressing the CD4 receptors and creates infection in cell-mediated immunity. Human immune deficiency virus type-1 (HIV-1) is the causative agent of acquired immune deficiency syndrome (AIDS) and AIDS-related complex. The available antiretroviral drugs for the treatment of HIV infections are not very effective and therefore, there is always a demand for new anti-HIV agents with lesser adverse effects while maintaining the activity against HIV mutants. Many enzymatic reactions are involved in the replication of this virus. Reverse transcriptase (RT) is a multifunctional heterodimeric enzyme that converts HIV RNA into proviral DNA. It is a multifunctional enzyme which is crucial to viral replication. Nucleoside RT inhibitors and non-nucleoside RT inhibitors targeting HIV-1 reverse transcriptase are two classes of drugs that are being used clinically to treat the HIV infection and AIDS. Non-nucleoside analog reverse transcriptase inhibitors inhibit viral DNA replication by binding at the allosteric non-bonding site of RT, causing a conformational change of the active site, all non-nucleoside RT inhibitors have unique antiviral potency with generally low toxicity and favorable pharmacokinetic properties because of wide range of chemically diverse structures and have important place in clinical use.

Several techniques are used in the design and development of HIV inhibitors. Quantitative structure–activity relation (QSAR) models can reveal significant correlations between biological activity and physicochemical parameters and can be used to improve the structure of inhibitor molecules and to interpret the improved structure in terms of favorable biological interaction. QSAR analysis has been widely used to modify lead compounds and it provides better rationale to understand the interaction mechanisms between chemical compounds and biological targets. Therefore, the development of drugs with significantly improved resistance profiles for chronic use in anti-HIV combination therapy is a major challenge.

With this goal in mind, the quantitative structure activity relationship (QSAR) analysis was performed on a series of aryluracil derivative analogues to
identify the physicochemical properties that govern the inhibitory activity of aryluracil derivative analogues. It is envisioned that these studies will produce models that can be used for future designing of new analogues with higher potency. In the present series, the aim is to correlate the physicochemical and structural requirements of these compounds to exhibit optimal inhibitory potency of NNRTIs enzyme which will in turn help in the modeling of HIV-1 NNRTIs inhibitors. For this purpose we have taken the activity data (IC50) that were reported by Xiaowei Wang, et al. The QSAR results reveal that the reverse transcriptase inhibitors activity could be modeled using different DFT-based descriptors such as softness (S), hardness (η), chemical potential (µ) and lowest unoccupied molecular orbital energy (LUMO) and empirical descriptors such as molecular weight (MW), surface tension (ST), index of refraction (IOR) and equalized electro-negativity (Xeq).

Computational Details

Two dimensional (2D) structures of various derivatives of substituted aryluracils (I) derivatives were drawn using ACD Lab Chem Sketch version 12.0 software and physicochemical and hydrophobic parameter such as molecular weight (Mw), molecular volume (Mv), molar refractivity (Mr), parachor (Pc), index of refraction (IOR), surface tension (st), density (D), polarizability (Pz) and partition coefficient (LogP) were calculated with the help of this software. The topological parameters such as balaban indices (J), Wiener index (W), mean wiener index (WA), Balaban centric index (BAC) and molecular connectivity (χ) were calculated by using E-Dragon software. 3D structures of compounds were also drawn on Accelrys’s Discovery Studio 3.5 program for calculation of various quantum chemical descriptors, such as total energy (TE), softness (S), hardness (η), chemical potential (µ), binding or cohesive energy (BE), band gap (BG), highest occupied molecular orbital energy (HOMO) and lowest unoccupied molecular orbital energy (LUMO). All species were fully optimized by density-functional theory at GGA-PW91 level in gas phase. Multiple regression analysis of the data gave several regression models. Regression analysis was performed by using SPSS 7.5 version. In the present work QSAR analysis were performed on various derivatives of aryluracils to obtain mathematical equations for correlating the structural variables and activity by applying Hansch analysis.

Results and Discussion

QSAR studies were performed on a set of 16 compounds of 1-[2-benzylalkoxy]methyl]-6-halo-6-aryluracils derivatives. On the basis of diversity between different substituent and reported biological activities, this series of compounds have been selected for QSAR analysis. In this series HIV-1 inhibitory activity has been expressed as IC50 reported by Xiaowei Wang et al. (2012) in micro-molar (µm) units which were converted to their log units (log IC50) and used in the present investigation. This was done in order to reduce skewness of the data set. IC50 represents the concentration of drug that inhibits 50% of NNRTIs enzyme. The biological activity data of these compounds were correlated with different physicochemical and steric parameters such as molecular weight (Mw), molecular volume (Mv), molar refractivity (Mr), parachor (Pc), polarizability (Pz) molecular connectivity (χ1 to χ5), topological parameters such as Wiener index (W), Balaban connectivity distance index (J), and DFT parameters such as BE, BG, S, η, µ, HOMO and LUMO, which have been found to be useful in QSAR-based drug modeling. Structural details of compounds with their experimental activity (log IC50) are given in Table I and in order to study the role of different substituent at different positions, indicator parameters as I1 for Iodine (-I) atom and I2 for Bromine (-Br) atom at R2 position were introduced, are also listed in Table I. In Table I, if a given group is present at the particular position then we assign it as 1 however, if substituent is not present at a given position then we consider it as zero.

The value of physicochemical, topological and quantum chemical parameters, which are used for making various significant models, are given in Table II and statistically significant models were obtained.
Compd | R_1 | R_2 | R_3 | I_1 | I_2 | pIC_{50} 
--- | --- | --- | --- | --- | --- | --- 
1. | CH_3 | I | H | 1 | 0 | 6.943 
2. | Ph | I | H | 1 | 0 | 8.000 
3. | CH_3 | Br | H | 0 | 1 | 5.739 
4. | Ph | Br | H | 0 | 1 | 5.818 
5. | CH_3 N(CH_3)_2 | H | 0 | 0 | 6.791 
6. | Ph N(CH_3)_2 | H | 0 | 0 | 7.921 
7. | CH_3 | I | CH_3 | 1 | 0 | 6.959 
8. | Ph | I | CH_3 | 1 | 0 | 8.523 
9. | CH_3 | Br | CH_3 | 0 | 1 | 6.678 
10. | Ph | Br | CH_3 | 0 | 1 | 7.367 
11. | CH_3 N(CH_3)_2 | CH_3 | 0 | 0 | 7.409 
12. | Ph N(CH_3)_2 | CH_3 | 0 | 0 | 7.678 
13. | CH_3 | I | F | 1 | 0 | 4.926 
14. | Ph | I | F | 1 | 0 | 7.824 
15. | CH_3 C_6H_5 | H | 0 | 0 | 6.824 
16. | Ph C_6H_5 | H | 0 | 0 | 7.699 

when some of these parameters were combined with the indicator parameter.

The biological activity data were correlated with different molecular descriptors for developing various significant models. Before obtaining a statistically significant model one has to examine whether or not any collinearity exists between the parameters used. This is done by obtaining correlation matrix, which is obtained in the present case with the physicochemical and topological parameters and also indicator parameter along with activity and is shown in Table III, correlation between 3D parameters and indicator parameter is given in Table IV and also the correlation between physicochemical and 3D parameters with indicator parameter is given in Table V. Multiple regression analysis using SPSS 7.5 version of the data gave several regression models.

The mono-parametric models cannot be used for modeling of the pIC_{50} because the quality of statistical data is not very good and the same thing applies to bi-parametric models which were also discarded. Hence, an attempt has been made to obtain multi-parametric models. Statistically significant multiparametric models were obtained when one of the physicochemical and topological parameters such as (Mv), (Mr), (Pc), (Pz), (LogP), (W), (Bac), (\(\chi\)) and DFT parameters such as BE, BG, S, \(\eta\), \(\mu\), HOMO and LUMO is combined with the indicator parameter. Multiple regression analysis resulted in several significant QSAR models.

Table II — Different descriptors value calculated from E-dragon and discovery softwares

<table>
<thead>
<tr>
<th>S.no</th>
<th>Mr</th>
<th>Mv</th>
<th>Pc</th>
<th>Pz</th>
<th>(\eta)</th>
<th>(\mu)</th>
<th>TE</th>
<th>HOMO LUMO</th>
<th>DIPOLE</th>
<th>((\eta))</th>
<th>(s)</th>
<th>((\mu))</th>
<th>BE</th>
<th>BG</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>83.640</td>
<td>230.000</td>
<td>635.800</td>
<td>33.150</td>
<td>6.819</td>
<td>5.345</td>
<td>4.309</td>
<td>-7.798.7</td>
<td>-0.215</td>
<td>0.092</td>
<td>2.170</td>
<td>0.0615</td>
<td>0.0308</td>
<td>-0.154</td>
</tr>
<tr>
<td>2</td>
<td>103.720</td>
<td>272.900</td>
<td>769.600</td>
<td>41.120</td>
<td>8.802</td>
<td>7.024</td>
<td>5.601</td>
<td>-7.990.5</td>
<td>-0.215</td>
<td>0.093</td>
<td>2.092</td>
<td>0.061</td>
<td>0.0305</td>
<td>-0.154</td>
</tr>
<tr>
<td>3</td>
<td>77.870</td>
<td>228.900</td>
<td>612.900</td>
<td>30.870</td>
<td>6.819</td>
<td>5.345</td>
<td>4.309</td>
<td>-3.452.1</td>
<td>-0.218</td>
<td>0.090</td>
<td>2.114</td>
<td>0.064</td>
<td>0.032</td>
<td>-0.154</td>
</tr>
<tr>
<td>4</td>
<td>97.730</td>
<td>273.100</td>
<td>746.700</td>
<td>38.740</td>
<td>8.802</td>
<td>7.024</td>
<td>5.601</td>
<td>-3.643.8</td>
<td>-0.219</td>
<td>0.094</td>
<td>1.692</td>
<td>0.063</td>
<td>0.031</td>
<td>-0.157</td>
</tr>
<tr>
<td>5</td>
<td>83.850</td>
<td>249.000</td>
<td>666.500</td>
<td>33.240</td>
<td>7.167</td>
<td>6.277</td>
<td>4.616</td>
<td>-1.012.3</td>
<td>-0.184</td>
<td>0.077</td>
<td>1.795</td>
<td>0.0535</td>
<td>0.02675</td>
<td>-0.130</td>
</tr>
<tr>
<td>6</td>
<td>103.930</td>
<td>291.800</td>
<td>800.300</td>
<td>41.200</td>
<td>9.150</td>
<td>7.956</td>
<td>5.908</td>
<td>-1.124.0</td>
<td>-0.184</td>
<td>0.078</td>
<td>1.713</td>
<td>0.052</td>
<td>0.026</td>
<td>-0.131</td>
</tr>
<tr>
<td>7</td>
<td>92.880</td>
<td>261.600</td>
<td>712.300</td>
<td>36.820</td>
<td>7.370</td>
<td>6.229</td>
<td>4.534</td>
<td>-0.787.4</td>
<td>-0.213</td>
<td>0.098</td>
<td>2.343</td>
<td>0.075</td>
<td>0.02875</td>
<td>-0.155</td>
</tr>
<tr>
<td>8</td>
<td>112.970</td>
<td>304.300</td>
<td>846.100</td>
<td>44.780</td>
<td>9.354</td>
<td>7.907</td>
<td>5.826</td>
<td>-0.799.0</td>
<td>-0.213</td>
<td>0.091</td>
<td>2.216</td>
<td>0.061</td>
<td>0.0305</td>
<td>-0.152</td>
</tr>
<tr>
<td>9</td>
<td>82.520</td>
<td>261.400</td>
<td>899.400</td>
<td>34.690</td>
<td>8.259</td>
<td>6.819</td>
<td>5.345</td>
<td>-0.353.07</td>
<td>-0.216</td>
<td>0.094</td>
<td>1.692</td>
<td>0.064</td>
<td>0.032</td>
<td>-0.152</td>
</tr>
<tr>
<td>10</td>
<td>107.380</td>
<td>305.600</td>
<td>823.200</td>
<td>42.560</td>
<td>8.802</td>
<td>7.024</td>
<td>5.601</td>
<td>-0.372.24</td>
<td>-0.213</td>
<td>0.087</td>
<td>1.559</td>
<td>0.063</td>
<td>0.0315</td>
<td>-0.150</td>
</tr>
<tr>
<td>11</td>
<td>93.100</td>
<td>280.500</td>
<td>743.000</td>
<td>36.900</td>
<td>7.718</td>
<td>7.161</td>
<td>4.841</td>
<td>-1.109.9</td>
<td>-0.182</td>
<td>0.078</td>
<td>1.973</td>
<td>0.035</td>
<td>0.02675</td>
<td>-1.285</td>
</tr>
<tr>
<td>12</td>
<td>113.180</td>
<td>323.200</td>
<td>876.800</td>
<td>44.860</td>
<td>9.702</td>
<td>8.839</td>
<td>6.133</td>
<td>-1.282.6</td>
<td>-0.182</td>
<td>0.076</td>
<td>1.886</td>
<td>0.053</td>
<td>0.0265</td>
<td>-0.129</td>
</tr>
<tr>
<td>13</td>
<td>115.120</td>
<td>316.900</td>
<td>834.300</td>
<td>45.640</td>
<td>7.370</td>
<td>6.229</td>
<td>4.434</td>
<td>-0.799.37</td>
<td>-0.222</td>
<td>0.099</td>
<td>1.7398</td>
<td>0.0615</td>
<td>0.0308</td>
<td>-0.160</td>
</tr>
<tr>
<td>14</td>
<td>110.030</td>
<td>311.800</td>
<td>815.100</td>
<td>43.620</td>
<td>9.354</td>
<td>7.907</td>
<td>5.826</td>
<td>-0.819.0</td>
<td>-0.221</td>
<td>0.099</td>
<td>1.737</td>
<td>0.061</td>
<td>0.0305</td>
<td>-0.152</td>
</tr>
<tr>
<td>15</td>
<td>103.950</td>
<td>281.900</td>
<td>784.300</td>
<td>41.200</td>
<td>7.167</td>
<td>6.277</td>
<td>4.616</td>
<td>-0.960.5</td>
<td>-0.205</td>
<td>0.072</td>
<td>1.970</td>
<td>0.0665</td>
<td>0.03325</td>
<td>-0.138</td>
</tr>
<tr>
<td>16</td>
<td>83.860</td>
<td>239.100</td>
<td>650.500</td>
<td>33.240</td>
<td>9.150</td>
<td>7.956</td>
<td>5.908</td>
<td>-1.155.8</td>
<td>-0.204</td>
<td>0.071</td>
<td>2.034</td>
<td>0.0665</td>
<td>0.03325</td>
<td>-0.137</td>
</tr>
</tbody>
</table>
Table V — Correlation matrix demonstrating physicochemical and DFT parameters with indicator parameters

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>Mr</th>
<th>Mv</th>
<th>Pc</th>
<th>Pz</th>
<th>Vχ1</th>
<th>Vχ2</th>
<th>Vχ3</th>
<th>I1</th>
<th>I2</th>
<th>I3</th>
<th>I4</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mr</td>
<td>0.723</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mv</td>
<td>0.630</td>
<td>0.907</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pc</td>
<td>0.699</td>
<td>0.983</td>
<td>0.967</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pz</td>
<td>0.723</td>
<td>1.000</td>
<td>0.907</td>
<td>0.983</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vχ1</td>
<td>0.499</td>
<td>0.859</td>
<td>0.855</td>
<td>0.876</td>
<td>0.859</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vχ2</td>
<td>0.551</td>
<td>0.818</td>
<td>0.908</td>
<td>0.883</td>
<td>0.818</td>
<td>0.930</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vχ3</td>
<td>0.444</td>
<td>0.787</td>
<td>0.815</td>
<td>0.815</td>
<td>0.787</td>
<td>0.990</td>
<td>0.918</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I1</td>
<td>-0.637</td>
<td>-0.219</td>
<td>-0.197</td>
<td>-0.224</td>
<td>-0.219</td>
<td>0.004</td>
<td>-0.216</td>
<td>0.039</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I2</td>
<td>0.434</td>
<td>0.180</td>
<td>-0.156</td>
<td>0.029</td>
<td>0.180</td>
<td>-0.161</td>
<td>-0.258</td>
<td>-0.255</td>
<td>-0.426</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A = activity, Mr = Molar refractivity, Mv = Molar volume, Pc = Parachor, χ = Molecular connectivity index, I1, I2 = indicator parameter

Table IV — Correlation matrix demonstrating DFT parameters with indicator parameters

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>TE</th>
<th>HOMO</th>
<th>LUMO</th>
<th>Di</th>
<th>η</th>
<th>S</th>
<th>µ</th>
<th>BE</th>
<th>BG</th>
<th>I1</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TE</td>
<td>-0.259</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOMO</td>
<td>0.214</td>
<td>0.700</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LUMO</td>
<td>0.125</td>
<td>0.851</td>
<td>0.780</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Di</td>
<td>0.057</td>
<td>-0.419</td>
<td>-0.173</td>
<td>-0.130</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>η</td>
<td>-0.200</td>
<td>-0.287</td>
<td>-0.741</td>
<td>-0.320</td>
<td>0.033</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>-0.227</td>
<td>-0.212</td>
<td>-0.771</td>
<td>-0.217</td>
<td>0.108</td>
<td>0.832</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>µ</td>
<td>0.235</td>
<td>0.786</td>
<td>0.957</td>
<td>0.889</td>
<td>-0.238</td>
<td>-0.617</td>
<td>-0.591</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BE</td>
<td>-0.580</td>
<td>-0.425</td>
<td>-0.517</td>
<td>-0.574</td>
<td>0.304</td>
<td>0.213</td>
<td>0.250</td>
<td>-0.586</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BG</td>
<td>-0.240</td>
<td>-0.287</td>
<td>-0.812</td>
<td>-0.297</td>
<td>0.195</td>
<td>0.835</td>
<td>0.976</td>
<td>-0.661</td>
<td>0.279</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>I1</td>
<td>0.434</td>
<td>-0.941</td>
<td>-0.484</td>
<td>-0.655</td>
<td>0.534</td>
<td>0.143</td>
<td>0.061</td>
<td>-0.562</td>
<td>0.312</td>
<td>0.138</td>
<td>1.000</td>
</tr>
</tbody>
</table>

A = activity, TE = Total energy, HOMO = Highest occupied molecular orbital Energy, LUMO = Lowest unoccupied molecular orbital energy, Di = Dipole moment, η = Hardness, S = Softness, µ = Chemical potential, BE = Binding energy, BG = Band gap, I1 = Indicator parameter

Table V — Correlation matrix demonstrating physicochemical and DFT parameters with indicator parameters

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>Mr</th>
<th>I1</th>
<th>LUMO</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I1</td>
<td>0.434</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mr</td>
<td>0.723</td>
<td>0.180</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>LUMO</td>
<td>0.125</td>
<td>-0.655</td>
<td>0.016</td>
<td>1.000</td>
</tr>
</tbody>
</table>

A = activity, Mr = Molar refractivity, LUMO = Lowest unoccupied molecular orbital energy

Modeling of pIC50 using empirical parameters

pIC<sub>50</sub> = 0.055(±0.033) Mr
-0.132 (±0.821) I1
-0.696(±0.929) I2 + 2.010

n =16, R=0.783, R<sup>2</sup>=0.613, R<sup>2</sup><sub>A</sub>=0.516, S.E=0.653, F<sub>(3-12)</sub>= 6.325, Q= 1.199 .... (1)

The above QSAR model explains only 61.3% variance in the HIV-1 inhibitory activity which is not significant for designing the more efficient HIV inhibitor. But on checking the series thoroughly, it also observed that anti-HIV-1 activity of the compound no. 13 is very low against NNRTI-resistant and therefore it may safely be treated as an outlier. By taking compound no.13 as outlier, a drastic change was observed in R<sup>2</sup> value and model equation shown in Eq. 2.

pIC<sub>50</sub> = 0.041(±0.022) Mr
+0.214(±0.551) I1
-0.758(±0.592) I2 + 3.380

n=15, R=0.882, R<sup>2</sup>=0.778, R<sup>2</sup><sub>A</sub>=0.717, S.E=0.412, F<sub>(3-11)</sub>= 12.852, Q= 2.141 .... (2)

Eq. 2 explains 77.8% variance. It confer that the coefficient of molar refractivity (Mr) contributes positively towards the inhibitor activity of HIV-1.

Other significant models obtained in the present analysis by using physicochemical parameter after treating compound 13 as outliers and models are:

pIC<sub>50</sub> = 0.006 (±0.003) Pc
+0.393(±0.547) I1
-0.671(±0.598) I2 + 2.834

n=15, R=0.884, R<sup>2</sup>=0.781, R<sup>2</sup><sub>A</sub>=0.721, S.E=0.409, F<sub>(3-11)</sub>= 13.048, Q= 2.161 .... (3)

pIC<sub>50</sub> = 0.017(±0.009) Mv
+0.603(±0.568)I1
-0.597(±0.612) I2+2.427
The calculated model Eq. 2 to 8 indicate that the coefficient of different orders of molecular connectivity parameters (χ), molecular volume (Mv), molar refractivity (Mr), parachor (Pc) and polarizability (Pz) are positive and this indicates that bulkier substituents with more branching would have positive effect on inhibitory activity. With reference to indicator parameters, the positive sign of coefficients of indicator parameter I₁ shows that the Iodine (–I) atom at R₂ position has positive influence and the negative sign of coefficients of indicator parameter I₂ shows that the Bromine (–Br) atom R₂ position has a negative influence on inhibitor activity of HIV-1.

Modeling of pIC₅₀ using quantum chemical descriptors

\[
pIC_{50} = 11.319(±32.472) \text{ HOMO} \\
-0.585(±0.368) \text{ BE} \\
+0.977(±0.870) I₁ + 4.273
\]

n = 16, R = 0.876, R² = 0.618, R²₀ = 0.522, S.E = 0.649, F(3-12) = 6.463, Q = 1.211 .... (9)

With quantum chemical parameters, the obtained QSAR models explain 61.8% variance. After treating compound 13 as outliers, severe change has been observed in the above model and is being reported as follows

\[
pIC_{50} = +11.337(±20.546) \text{ HOMO} \\
-0.467(±0.240) \text{ BE} \\
+1.204(±5.61) I₁ + 5.278
\]

n = 15, R = 0.885, R² = 0.784, R²₀ = 0.725, S.E = 0.406, F(3-11) = 13.295, Q = 2.180 .... (10)

Eq. 10 explains 78.4% variance in the HIV inhibitory activity. The sign of HOMO is positive hence increase in the value of these parameters may enhance the activity of the drugs.

Other significant models obtained in the present analysis using quantum chemical parameter, after treating compound 13 as outliers are given below:

\[
pIC_{50} = +21.703(±29.394) \text{ µ} \\
-0.432(±0.242) \text{ BE} \\
+1.287(±0.566) I₁ + 6.355
\]

n = 15, R = 0.896, R² = 0.748, S.E = 0.388, F(3-11) = 14.881, Q = 2.309 .... (11)

\[
pIC_{50} = +21.751(±37.243) \text{ LUMO} \\
-0.450(±0.250) \text{ BE} \\
+1.312(±0.647) I₁ + 4.915
\]

n = 15, R = 0.887, R² = 0.729, S.E = 0.403, F(3-11) = 13.537, Q = 2.201 .... (12)

\[
pIC_{50} = +0.0001(±0.0001) \text{ TE} \\
-0.480(±0.240) \text{ BE} \\
+1.099(±0.530) I₁ + 3.409
\]

n = 15, R = 0.881, R² = 0.714, S.E = 0.414, F(3-11) = 12.435, Q = 2.108 .... (13)

\[
pIC_{50} = -0.316(±1.274) \text{ Di} \\
-0.516(±0.229) \text{ BE} \\
+1.156(±0.620) I₁ + 6.355
\]

n = 15, R = 0.873, R² = 0.696, S.E = 0.427, F(3-11) = 11.694, Q = 2.044 .... (16)
independent variable along with activity as a descriptor, we have taken both the descriptor as an indicator parameter, the positive sign of coefficients activity towards HIV-1 enzyme. In the view of indicator parameters, the positive sign of coefficients of the indicator parameter I₂ shows that the Iodine at R₂ position has positive influence towards inhibitor activity of HIV-1.

A close look of these models reveals that the out of several quantum chemical parameters, the coefficient of total energy (TE), chemical potential (µ), HOMO and LUMO are positive (Eq. 10-13) suggesting that increase in the value of these parameters should enhance the inhibitory activity towards HIV-1 enzyme of the compound. In Eq. 14-17 the coefficient of dipole, softness, hardness and band gap are negative suggesting that decrease in the value of these parameters may enhance the activity of the compound to inhibit HIV-1.

It is interesting to note that in all DFT based models 10-17, the coefficient of binding energy (BE) is negative in sign for all the obtained models and it indicates that on decreasing the value of this parameter, activity may enhance. With reference to indicator parameter, the positive coefficient of indicator parameter I₂ (Eq. 10 to 17) which represents the presence of Iodine (-I) at R₂ position suggests that Iodine (-I) at R₂ position is favorable for the activity and should be retained in future drugs of this series.

Out of several QSAR models, the Eq. 3 based on empirical and 11 using DFT based descriptor respectively show the best results. From Eq. 3 the compound having highest value of parachor (Pc) and in which Iodine (-I) is present at R₂ position while absence of Bromine (–Br) atom at R₂ position is favorable for inhibitory effect toward NNRTs enzyme. According to Eq. 11 increase in value of chemical potential (µ), and decrease in the value of binding energy (BE) and presence of Iodine (–I) at R₂ position may enhance the activity of compound toward NNRTs enzyme.

To see the correlation between empirical and DFT descriptor, we have taken both the descriptor as an independent variable along with activity as a dependent variable. The multiple linear regression analysis showed that the descriptor molar refractivity (Mr) with chemical potential (µ) and LUMO are statically related with activity of the compounds. The statically significant models are as:

\[ pIC_{50} = -23.61(±105.718) S -0.513(±0.233) BE +1.079(±0.543) I_1 + 3.322 \]

\[ n=15, \ R=0.872, \ R^2=0.760, \ R^2_A=0.695, \ S.E=0.428, \]

\[ F_{(3-11)} = 11.619, Q=2.037 \quad \ldots \ (17) \]

where,

\[ n = \text{total no. of compounds} \]

\[ R = \text{Correlation coefficient} \]

\[ R^2 = \text{Coefficient of determination} \]

\[ R^2_A = \text{Adjusted coefficient of determination} \]

\[ S.E. = \text{Standard error of estimate} \]

\[ F = \text{Variance ratio}^{15,16} \]

\[ Q = \text{Quality of fit}^{17,18} \]

Calculated model equations 18 & 19 showed that the coefficient of molar refraction (Mr) and the coefficient of DFT parameter µ & LUMO are positive. Positive coefficient indicates that if it increase then these parameters enhance the inhibitory activity towards HIV-1 enzyme. In the view of indicator parameters, the positive sign of coefficients of the indicator parameter I₂ shows that the Iodine at R₂ position has positive influence towards inhibitor activity of HIV-1.

In order to confirm that the model with excellent statistics also has excellent predictive power too, we have evaluated quality factor Q. The quality factor Q is the ratio of correlation coefficient to its standard error of estimation \( i.e. \ Q = R/S.E. \) thus higher the value of R, the lower the S.E., the greater will be the Q. The predictive power as determined by the Pogliani Q parameter for the model expressed by Eq. 3, 11 and 19 (Q = 2.161, 2.309 & 2.315) confirms that this model has statistically significant as well as excellent predictive power. The high values of R, R² and low values of S.E. indicate these model equations are statistically significant and the best among all tri-parametric models discussed above.

Predicted and residual activity values for Eqs. 3, 11 and 19 are given in Table VI respectively. Predicted values are the calculated activities of the equation and the residual values are the difference between the observed biological activities and the calculated activities and are found to be low. The calculated F value is greater than F theoretical value \( [F_{(3, 11)} = 3.59] \) for all the significant equations. The higher t test measures the statistical significance of the regression coefficients. It is pertinent to mention that \( R^2 \) goes
on increasing with each addition of descriptors, while SE goes on decreasing with successive addition of descriptors, and it means that both statistics as well as quality of the model goes on improving. The values of t test are given in Table VII.

The plot of observed pIC₅₀ versus predicted pIC₅₀ for Eq. 3, 11 and 19 are shown in graph (Figure 1, Figure 2 and Figure 3) and the predicted R² was found to be fairly large.

The cross validation analysis was performed using leave one out (LOO) method in which one compound is removed from the data set and the activity is correlated using the rest of the data set. The cross-validated R² in each case was found to be very close to the value of R² for the entire data set and hence these models can be termed as statistically significant. Cross validation provides the values of PRESS, SSY, PSE, R²CV and R²A from which we can test the predictive power of the proposed model are given in Table VIII.

The calculated cross-validated parameters confirm the validity of the models. Model no. 3, 11 and 19 fulfill all the requirements for ideal models, that is
The predictive error of coefficient of correlation (PE) is yet another parameter used to estimate the predictive power of the proposed models. It is argued that if the values $R < PE$, then such correlation is not significant; however if values are $R > PE$ by several times (at least three times), then values are correlated. However, if values are $R > 6PE$, then mathematically the correlation is undeniably good. For the models 3, 11 and 19 developed the condition $R > 6PE$ is satisfied and hence they can be said to have a good predictive power.

The outlier may be of particular interest, such as in the case of fraud detection, where may indicate fraudulent activity. Thus, outlier detection and analysis is an interesting data mining task, referred to as outlier analysis. In fact, an outlier is an observation that appears to deviate markedly from other observations in the sample and are leads to new understanding and covering them up by including them in QSAR, and at the cost of lower $R^2$.

## Conclusion

In this study we have performed different calculations based on empirical and DFT- descriptors to obtain the correlation of different descriptors with biological activity of a series of substituted aryluracils derivatives. The calculated QSAR results based on empirical parameters demonstrate that molecules with high parachor (Pc) value should be preferred for future modeling. As far as indicator parameters are concerned it may be said that the presence of Iodine at $R_2$ position may enhance the activity towards NNRTs.

The multiple linear regression (MLR) models were developed with quantum chemical descriptors including total energy (TE), softness ($S$), hardness ($\eta$), highest unoccupied molecular orbital energy (HOMO) and lowest unoccupied molecular orbital energy (LUMO) and it was found that statistical quality as demonstrated by $R^2$ and low prediction errors is pretty good.
good. The best MLR model based on DFT descriptors demonstrates that high increase in the value chemical potential ($\mu$) while decrease in the value of cohesive or binding energy (BE) and the presence of Iodine at $R_2$ position may be helpful for future designing.

**Acknowledgments**

The author A. K. Srivastava gratefully acknowledges support from the University Grants Commission (UGC) New Delhi, India through the grant [Project no. DST/40-70/2011 (SR)]. Author A. Singh wishes to acknowledge the UGC, New Delhi, India, for the Dr. D. S. Kothari’s postdoctoral fellowship.

**References**