3D QSAR Analysis on Arylbenzofuran derivatives as Histamine H₃ Receptor Inhibitors using k Nearest Neighbor Molecular Field Analysis (KNNMFA)

Sanmati K. Jain* and Sudip Nag

*SLT Institute of Pharmaceutical Sciences, Guru Ghasidas Vishwavidyalaya (Central University), Bilaspur (Chhattisgarh)-495009, India.

*Corres. author: sanmatijain72@yahoo.co.in
Phone No. 07752-260027, Mob: +919827088484

Abstract: Three dimensional quantitative structure activity relationship (3D QSAR) analysis using k nearest neighbor molecular field analysis (kNN MFA) method was performed on a series of arylbenzofuran derivatives as histamine h₃ receptor inhibitors using molecular design suite (VLifeMDS). This study was performed with 58 compounds (data set) using sphere exclusion (SE) algorithm and random selection method for the division of the data set into training and test set. kNN-MFA methodology with step wise variable selection forward-backward, Simulated Annealing and Genetic Algorithms methods were used for building the QSAR models. Two predictive models were generated with SW-kNN MFA. The most predictive model was generated by kNN (stepwise forward backward) using sphere exclusion data selection method. This model explains good internal (q² = 0.5445) as well as very good external (Predr² = 0.9013) predictive power of the model. The steric descriptors at the grid points, S_2307 and S_2377 plays important role in imparting activity. This model indicates that two steric descriptors are involved. The kNN-MFA contour plots provided further understanding of the relationship between structural features of substituted arylbenzofuran derivatives and their activities which should be applicable to design newer potential H₃ receptor inhibitors.

Keywords: 3D-QSAR, kNN-MFA, h₃ receptor inhibitors, arylbenzofuran derivatives.

INTRODUCTION

Histamine is a biogenic amine that influences a wide range of pathophysiological processes. At present, four subtypes of histamine G protein-coupled receptors (GPCRs) are known. H₁ and H₂ receptors are concerned in allergic responses and gastric acid secretion, respectively. The most recently discovered H₄ receptor is mainly located on mast cells, eosinophils and lymphoid tissues and seems to be involved in inflammatory processes (1). The histamine H₃ receptor was identified in 1983 and was initially described as an autoreceptor, mainly expressed in the central nervous system (CNS), regulating histamine biosynthesis and release from histaminergic neurons (2). Histamine H₃ receptors are expressed in the central nervous system and to a lesser extent the peripheral nervous system, where they act as autoreceptors in presynaptic histaminergic neurons, and also control histamine turnover by feedback inhibition of histamine synthesis and release (3). Subsequently, H₃ receptors have also been shown to act as heteroreceptors on non-histaminergic neurons, where they inhibit the release of other neurotransmitters such as acetylcholine, dopamine, norepinephrine, serotonin and various neuropeptides (4-9).

The high density of H₃ receptors in different CNS areas and their influence on the release of a large variety of neurotransmitters encouraged wide pharmacological investigation on their physiological role and the quest for potential
therapeutic applications of H₃-antagonists in the treatment of various CNS diseases. Among them, the most promising ones include attention-deficit hyperactivity disorders (ADHD), Alzheimer’s disease, epilepsy, schizophrenia, obesity and eating disorders (10-19).

Since the discovery of the reference antagonist thioperamide (20), many classes of potent and selective H3-antagonists have been reported. The earliest generation of H₃-antagonists was derived from the endogenous neurotransmitter histamine and the compounds contained an imidazole ring in their structures (21,22). The imidazole-based H₃ antagonists such as ciproxifan, thioperamide are useful tools and reference standards for pharmacological research, but they have two drawbacks, which includes:

i) Imidazole-based analogues have poor CNS penetration (23,24) while low CNS penetration may be a desirable property in a drug targeted to the treatment of peripheral diseases.

ii) Inhibit cytochrome-P₄₅₀ (CYP) drug metabolizing enzymes, leading to drug-drug interactions by inhibiting the metabolism of co-administered drugs (25-29).

iii) Additionally, drugs that inhibit CYP enzymes have the potential to alter the endogenous metabolism of important circulating hormones.

For this reason non-imidazoles have been targeted as potential drug candidates, and several distinct classes of non-imidazole H₃ antagonists (30-36) have been produced. The increasing interest in the therapeutic potential of H₃ antagonists are driving current medicinal chemistry efforts to identify potent, selective, therapeutically efficacious and safe agents for clinical development.

**Experimental**

**The Work Station:** Workstations are raster systems in which a computer with a full operating system and mass storage facility is integrated with graphical display. All the computational methods were performed on HCL PC using QSARPlus version 1.0.

**Data sets:** In the present study 58 molecules of arylbenzofuran derivatives were used which were reported to have Histamine H₃-receptor antagonist activity (38). The activity data [binding affinity (µM)] determined by using human and rat histamine H₃-receptor, activity data [pKi] have been converted to the logarithmic scale [pKi (moles)] and then used for subsequent QSAR analysis as the response variables. pKi data saved as .txt file. The various substituents of all compounds along with their actual biological activities are shown in Table 1.

**Methodology:** Structure of the molecule drawn in the 2DDraw app option in Tool menu of QSARPlus. Then 2D structures where exported to QSARPlus window (2D structure converted to 3D structure). After the conversion, structures are saved as .mol2 file in QSARPlus 3D window. Energy minimization is done by using Merck molecular force field (MMFF) method which results in the optimization of the geometry of the molecule using the following criteria.
<table>
<thead>
<tr>
<th>Force field</th>
<th>Charge</th>
<th>Max. no. of cycles</th>
<th>Convergence criteria (rms gradients)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMFF</td>
<td>MMFF</td>
<td>10,000</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Dielectric properties
- Distance dependent function
- Constant – 1.0
- Gradient type analysis – 1.0

Non-bonded cut off: Electro static – 20.00; VdW – 10.00.

Models were generated by k-NN-MFA in conjunction with stepwise (SW) forward-backward [Cross correlation – 0.5; Term selection – q2; Variance cut off – 0.0; No. of Max. Neighbours – 5; No. of Min. Neighbours – 2; Select prediction method – Distance based weighted average], Simulated Annealing [Maximum temperature-100.0; Minimum temperature-0.01; Decrease temperature by-10.0; Iteration at given temperature-5; Terms in model-4; Perturbation limit-5; Cross correlation limit-1.0; Term selection criteria-q2; Seed-0; Scaling-Autoscaling; No. of Max. Neighbours – 5; No. of Min. Neighbours – 2; Select prediction method – Distance based weighted average] and Genetic Algorithms [Cross correlation limit-1.0; Cross over probability-0.9; Mutation probability-0.1; Population-10; Number of generations-1000; Print after iterations-100; Convergence ending criteria-0; Chromosome length-3; Seed-0; Term selection criteria-q2; Scaling-Autoscaling; No. of Max. Neighbours – 5; No. of Min. Neighbours – 2; Select prediction method – Distance based weighted average] variable selection methods. Once the training and test sets were generated, kNN methodology was applied to the descriptors generated over the grid.

Molecular Alignment: In 3D QSAR first the alignment of the optimized molecules is done by using Template based alignment method.

Model development: The activity data were subjected to k-nearest neighbor molecular field analysis method for model building.

Data selection: Biological activity taken as dependent variable and descriptors as independent variable. Following methods were used for creation of training and test set.
- Random selection method
- Sphere Exclusion method

Random selection: In order to construct and validate the QSAR models, both internally and externally, the data sets were divided into training [85%-75% (85%, 80% and 75%) of total data set] and test sets [15%-25% (15%, 20% and 25%) of total data set] in a random manner. 05 trials were run in each case.

Sphere Exclusion method: In this method initially data set were divided into training and test set using sphere exclusion method. In this method dissimilarity value provides an idea to handle training and test set size. It needs to be adjusted by trial and error until a desired division of training and test set is achieved. Increase in dissimilarity value results in increase in number of molecules in the test set. After the creation of training and test set, Min and Max value of the test and training set is checked, using the QSAR tool, if the values are not following the Min – Max, then the training / test set is again set and procedure is repeated. If the Min – Max is following, then the k nearest neighbor molecular field analysis method was used for model building.
Table 1: General structure of the compounds of Arylbenzofuran derivatives and their biological activities (data set of 58 molecules)

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Compound</th>
<th>Benzofuran substituent</th>
<th>Phenyl Substituent</th>
<th>hH₃(nM)</th>
<th>log (1/hH₃)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>H</td>
<td>4'-CN</td>
<td>0.45</td>
<td>9.347</td>
</tr>
<tr>
<td>2</td>
<td>4a</td>
<td>H</td>
<td>3'-CN</td>
<td>0.27</td>
<td>9.569</td>
</tr>
<tr>
<td>3</td>
<td>4b</td>
<td>H</td>
<td>4'-F</td>
<td>3.2</td>
<td>8.495</td>
</tr>
<tr>
<td>4</td>
<td>4c</td>
<td>H</td>
<td>3'-F</td>
<td>2.2</td>
<td>8.658</td>
</tr>
<tr>
<td>5</td>
<td>4d</td>
<td>H</td>
<td>4'-Cl</td>
<td>6.3</td>
<td>8.201</td>
</tr>
<tr>
<td>6</td>
<td>4e</td>
<td>H</td>
<td>3'-Cl</td>
<td>5.0</td>
<td>8.301</td>
</tr>
<tr>
<td>7</td>
<td>4f</td>
<td>H</td>
<td>2'-Cl</td>
<td>5.3</td>
<td>8.276</td>
</tr>
<tr>
<td>8</td>
<td>4g</td>
<td>H</td>
<td>4'-CF₃</td>
<td>5.7</td>
<td>8.244</td>
</tr>
<tr>
<td>9</td>
<td>4h</td>
<td>H</td>
<td>3'-CF₃</td>
<td>3.9</td>
<td>8.409</td>
</tr>
<tr>
<td>10</td>
<td>4i</td>
<td>H</td>
<td>4'-Me</td>
<td>7.9</td>
<td>8.102</td>
</tr>
<tr>
<td>11</td>
<td>4j</td>
<td>H</td>
<td>3'-Me</td>
<td>2.7</td>
<td>8.569</td>
</tr>
<tr>
<td>12</td>
<td>4k</td>
<td>H</td>
<td>2'-Me</td>
<td>4.1</td>
<td>8.387</td>
</tr>
<tr>
<td>13</td>
<td>4l</td>
<td>H</td>
<td>4'-OCF₃</td>
<td>7.6</td>
<td>8.119</td>
</tr>
<tr>
<td>14</td>
<td>4m</td>
<td>H</td>
<td>3'-OCF₃</td>
<td>11</td>
<td>7.959</td>
</tr>
<tr>
<td>15</td>
<td>4n</td>
<td>H</td>
<td>4'-OMe</td>
<td>9.1</td>
<td>8.041</td>
</tr>
<tr>
<td>16</td>
<td>4o</td>
<td>H</td>
<td>3'-OMe</td>
<td>1.2</td>
<td>8.921</td>
</tr>
<tr>
<td>17</td>
<td>4p</td>
<td>H</td>
<td>2'-OMe</td>
<td>15</td>
<td>7.824</td>
</tr>
<tr>
<td>18</td>
<td>4q</td>
<td>H</td>
<td>3'-Cl,4'-Cl</td>
<td>3.6</td>
<td>8.444</td>
</tr>
<tr>
<td>19</td>
<td>4r</td>
<td>H</td>
<td>3'-Cl,5'-Cl</td>
<td>5.0</td>
<td>8.301</td>
</tr>
<tr>
<td>20</td>
<td>4s</td>
<td>H</td>
<td>3'-Me,4'-Me</td>
<td>5.9</td>
<td>8.229</td>
</tr>
<tr>
<td>21</td>
<td>4t</td>
<td>H</td>
<td>3'-Me,5'-Me</td>
<td>3.6</td>
<td>8.444</td>
</tr>
<tr>
<td>22</td>
<td>4u</td>
<td>H</td>
<td>4'-COOMe</td>
<td>1.9</td>
<td>8.721</td>
</tr>
<tr>
<td>23</td>
<td>4v</td>
<td>H</td>
<td>3'-C(O)Me</td>
<td>0.084</td>
<td>10.076</td>
</tr>
<tr>
<td>24</td>
<td>4w</td>
<td>H</td>
<td>4'-CH₂OH</td>
<td>0.88</td>
<td>9.056</td>
</tr>
<tr>
<td>25</td>
<td>4x</td>
<td>H</td>
<td>3'-CH₂OH</td>
<td>0.49</td>
<td>9.310</td>
</tr>
<tr>
<td>26</td>
<td>8a</td>
<td>H</td>
<td>4'-Br</td>
<td>7.7</td>
<td>8.114</td>
</tr>
<tr>
<td>27</td>
<td>8b</td>
<td>H</td>
<td>4'-CN,2'-Me</td>
<td>2.0</td>
<td>8.699</td>
</tr>
<tr>
<td>28</td>
<td>8c</td>
<td>H</td>
<td>4'-CN,3'-Me</td>
<td>0.28</td>
<td>9.553</td>
</tr>
<tr>
<td>29</td>
<td>8d</td>
<td>H</td>
<td>4'-CN,3'-F</td>
<td>0.64</td>
<td>9.194</td>
</tr>
<tr>
<td>30</td>
<td>8e</td>
<td>7-F</td>
<td>4'-CN</td>
<td>0.43</td>
<td>9.367</td>
</tr>
<tr>
<td>31</td>
<td>8f</td>
<td>7-Me</td>
<td>4'-CN</td>
<td>1.1</td>
<td>8.959</td>
</tr>
<tr>
<td>32</td>
<td>8g</td>
<td>6-Me</td>
<td>4'-CN</td>
<td>0.77</td>
<td>9.114</td>
</tr>
<tr>
<td>33</td>
<td>9</td>
<td>H</td>
<td>3'-CH(OH)Me</td>
<td>0.44</td>
<td>9.357</td>
</tr>
<tr>
<td>34</td>
<td>10</td>
<td>H</td>
<td>3'-C(OH)Me₂</td>
<td>0.25</td>
<td>9.602</td>
</tr>
<tr>
<td>35</td>
<td>11</td>
<td>H</td>
<td>3'-COOH</td>
<td>15</td>
<td>7.824</td>
</tr>
<tr>
<td>36</td>
<td>12</td>
<td>H</td>
<td>3'-C(O)N(Me)OMe</td>
<td>0.62</td>
<td>9.208</td>
</tr>
<tr>
<td>37</td>
<td>13a</td>
<td>H</td>
<td>3'-C(O)Et</td>
<td>0.23</td>
<td>9.638</td>
</tr>
<tr>
<td>38</td>
<td>13b</td>
<td>H</td>
<td>3'-C(O)CH₂CHMe₂</td>
<td>0.68</td>
<td>9.168</td>
</tr>
<tr>
<td>39</td>
<td>13c</td>
<td>H</td>
<td>3'-C(O)-(3'-F)C₆H₄</td>
<td>2.6</td>
<td>8.585</td>
</tr>
<tr>
<td>40</td>
<td>13d</td>
<td>H</td>
<td>3'-CHO</td>
<td>0.40</td>
<td>9.398</td>
</tr>
<tr>
<td>41</td>
<td>14a</td>
<td>H</td>
<td>3'-C(=NOH)Me</td>
<td>0.49</td>
<td>9.310</td>
</tr>
<tr>
<td>42</td>
<td>14b</td>
<td>H</td>
<td>3'-(=NOMe)Me</td>
<td>0.42</td>
<td>9.377</td>
</tr>
<tr>
<td>43</td>
<td>14c</td>
<td>H</td>
<td>3'-C(=NOEt)Me</td>
<td>3.2</td>
<td>8.495</td>
</tr>
<tr>
<td>44</td>
<td>14d</td>
<td>H</td>
<td>3'-C(=NO+Bu)Me</td>
<td>4.3</td>
<td>8.367</td>
</tr>
<tr>
<td>45</td>
<td>15a</td>
<td>H</td>
<td>4'-C(O)-C-Pr</td>
<td>0.26</td>
<td>9.585</td>
</tr>
<tr>
<td>46</td>
<td>15b</td>
<td>H</td>
<td>3'-C(O)-C-Pr</td>
<td>0.21</td>
<td>9.678</td>
</tr>
</tbody>
</table>
RESULTS AND DISCUSSION
Following statistical parameters were used to correlate biological activity and molecular descriptors: n = number of molecules, Vn = number of descriptors, k = number of nearest neighbor, df = degree of freedom, r² = coefficient of determination, q² = cross validated r² (by the leave-one out method), pred_r² = r² for external test set, pred_r²se = coefficient of correlation of predicted data set. Selecting training and test set by random selection method, the Unicolumn statics was performed which is shown in Table 2. The unicolumn statistical analysis can be interpreted that the mean and standard deviation for the training and test sets provide insight into the relative difference in the mean and point density distribution of the two sets. The minimum and maximum values in both the training and test sets are compared such that the maximum of the test set should be less than that of the training set. The minimum of the test set should be greater than that of the training set, suggesting the interpolative behavior of the test set (i.e., derived within the minimum–maximum range of the training set) which is prerequisite analysis for further QSAR study.

<table>
<thead>
<tr>
<th>No.</th>
<th>No.</th>
<th>Substituent</th>
<th>4'-CN</th>
<th>r²</th>
<th>q²</th>
</tr>
</thead>
<tbody>
<tr>
<td>47</td>
<td>16</td>
<td>3-I</td>
<td>4'-CN</td>
<td>0.51</td>
<td>9.292</td>
</tr>
<tr>
<td>48</td>
<td>17a</td>
<td>3-Cl</td>
<td>4'-CN</td>
<td>0.39</td>
<td>9.409</td>
</tr>
<tr>
<td>49</td>
<td>17b</td>
<td>3-Cl,6-Cl</td>
<td>4'-CN</td>
<td>0.83</td>
<td>9.081</td>
</tr>
<tr>
<td>50</td>
<td>18a</td>
<td>3-Br</td>
<td>4'-CN</td>
<td>0.29</td>
<td>9.538</td>
</tr>
<tr>
<td>51</td>
<td>18b</td>
<td>3-Br</td>
<td>4'-CN, 3'-Me</td>
<td>1.3</td>
<td>8.886</td>
</tr>
<tr>
<td>52</td>
<td>19a</td>
<td>3-Ph</td>
<td>4'-CN</td>
<td>21</td>
<td>7.678</td>
</tr>
<tr>
<td>53</td>
<td>19b</td>
<td>3-(3',5'-DiMeC₆H₄)</td>
<td>4'-CN</td>
<td>60</td>
<td>7.222</td>
</tr>
<tr>
<td>54</td>
<td>19c</td>
<td>3-(3'-Pyridyl)</td>
<td>4'-CN</td>
<td>13</td>
<td>7.886</td>
</tr>
<tr>
<td>55</td>
<td>19d</td>
<td>3-(2'-Furyl)</td>
<td>4'-CN</td>
<td>2.8</td>
<td>8.553</td>
</tr>
<tr>
<td>56</td>
<td>19e</td>
<td>3-(3'-Thienyl)</td>
<td>4'-CN</td>
<td>2.1</td>
<td>8.678</td>
</tr>
<tr>
<td>57</td>
<td>19f</td>
<td>3-(3'''(2''CHO)thienyl)</td>
<td>4'-CN</td>
<td>0.73</td>
<td>9.137</td>
</tr>
<tr>
<td>58</td>
<td>20</td>
<td>3-(3''(2''CH₂OH)thienyl)</td>
<td>4'-CN</td>
<td>1.9</td>
<td>8.721</td>
</tr>
</tbody>
</table>

*Total data set of 58 molecules was used for the present study.

Figure 1: Training and test set data fitness plot for Model-1 (trial-1)
Data fitness plot for training and test set is shown in Figure 1. kNN-MFA plot is shown in Figure 2 and 3D alignment of the molecules is shown in Figure 3. Results of 3D-QSAR kNN-MFA analysis using sphere exclusion and random data selection methods are shown in Table-2 and Table 4-6 respectively. Result of the most predictive model generated by kNN (Trial-1, stepwise forward backward) method using sphere exclusion data selection method is as follows:

**kNN Method**

Training Set Size = 53  
Test Set Size = 5

**Selected Descriptors:**

S_{2307}  
S_{2377}

**Statistics:**

k Nearest Neighbour= 2  
n = 53  
Degree of freedom = 50  
$q^2 = 0.5445$  
$q^2_{se} = 0.4220$  
$Predr^2 = 0.9013$  
$pred_r^2se = 0.1955$

**Descriptor Range:**

S_{2307} -0.007384 -0.007189  
S_{2377} -0.00308 -0.002385

This model explains internal ($q^2 = 0.5445$) as well as very good external ($Predr^2 = 0.9013$) predictive power of the model. The steric descriptors at the grid points S_{2307} and S_{2377} plays important role in imparting biological activity. This model indicates that two steric descriptors are involved suggesting that activity is dominated by steric interactions. kNN-MFA result plot in which 3D-alignment of molecules with the important steric points contributing with model with ranges of values shown in parenthesis represented in Figure 2.

Finally, it is hoped that the work presented here will play an important role in understanding the relationship of physiochemical parameters with structure and biological activity. By studying the QSAR model one can select the suitable substituent for active compound with maximum potency.
Table 2: Results of 3D-QSAR analysis using kNN method (sphere exclusion)

<table>
<thead>
<tr>
<th>TRIAL</th>
<th>DISSIMILARITY VALUE*</th>
<th>TEST SET</th>
<th>kNN result</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Stepwise forward-backward</td>
</tr>
</tbody>
</table>
| 1     | 6.50                 | 4f,4q,8a,4t,4l | k Nearest Neighbour= 2  
n = 53  
Degree of freedom = 50  
q2 = \textbf{0.5445}  
q2_se = 0.4220  
Predr2 = \textbf{0.9013}  
pred_r2se = 0.1955 | k Nearest Neighbour= 4  
n = 53  
Degree of freedom = 48  
q2 = 0.0894  
q2_se = 0.5967  
Predr2 = -0.0971  
pred_r2se = 0.6520 | k Nearest Neighbour= 4  
n = 53  
Degree of freedom = 49  
q2 = 0.0816  
q2_se = 0.5993  
Predr2 = 0.4961  
pred_r2se = 0.4419 |
| 2     | 7.00                 | 4f,4q,8a,18b,4t,4x,4l | k Nearest Neighbour= 2  
n = 51  
Degree of freedom = 47  
q2 = 0.5601  
q2_se = 0.4203  
Predr2 = 0.4191  
pred_r2se = 0.4119 | k Nearest Neighbour= 3  
n = 51  
Degree of freedom = 46  
q2 = 0.1898  
q2_se = 0.5704  
Predr2 = 0.3218  
pred_r2se = 0.4451 | k Nearest Neighbour= 4  
n = 51  
Degree of freedom = 47  
q2 = 0.0627  
q2_se = 0.6135  
Predr2 = 0.5667  
pred_r2se = 0.3558 |
| 3     | 8.00                 | 4d,4f,4o,4q,8a,4t,4x,4g,8c | k Nearest Neighbour= 2  
n = 49  
Degree of freedom = 47  
q2 = 0.3541  
q2_se = 0.5096  
Predr2 = -0.3366  
pred_r2se = 0.6471 | k Nearest Neighbour= 4  
n = 49  
Degree of freedom = 44  
q2 = 0.3644  
q2_se = 0.5055  
Predr2 = 0.4201  
pred_r2se = 0.4262 | k Nearest Neighbour= 5  
n = 49  
Degree of freedom = 45  
q2 = 0.0799  
q2_se = 0.6083  
Predr2 = 0.1339  
pred_r2se = 0.5209 |
| 4     | 8.50                 | 4c,4d,4o,8a,4n,4n,4f,4x,4g,8d,8c,4t,19a | k Nearest Neighbour= 2  
n = 46  
Degree of freedom = 43  
q2 = 0.4779  
q2_se = 0.4496  
Predr2 = 0.1840  
pred_r2se = 0.5766 | k Nearest Neighbour= 2  
n = 46  
Degree of freedom = 41  
q2 = 0.1198  
q2_se = 0.5837  
Predr2 = -0.1785  
pred_r2se = 0.6929 | k Nearest Neighbour= 3  
n = 46  
Degree of freedom = 42  
q2 = 0.1329  
q2_se = 0.5794  
Predr2 = 0.3848  
pred_r2se = 0.5006 |
| 5     | 9.00                 | 4e,4d,4o,8a,18a,4n,8f,16,4x,4m,8d,8c,8g,4t,4c,2,41,19c | k Nearest Neighbour= 2  
n = 40  
Degree of freedom = 37  
q2 = 0.4292  
q2_se = 0.4808  
Predr2 = 0.0211  
pred_r2se = 0.5876 | k Nearest Neighbour= 4  
n = 40  
Degree of freedom = 35  
q2 = 0.1759  
q2_se = 0.5777  
Predr2 = -0.3135  
pred_r2se = 0.6806 | k Nearest Neighbour= 5  
n = 40  
Degree of freedom = 36  
q2 = 0.0383  
q2_se = 0.6241  
Predr2 = 0.2357  
pred_r2se = 0.5192 |

*Sphere exclusion data selection method was used in this study
Table 3: Min-Max table for model (Trial-1, stepwise forward backward)

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Average</th>
<th>Max</th>
<th>Min</th>
<th>StdDev</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set</td>
<td>8.8166</td>
<td>10.0760</td>
<td>7.2220</td>
<td>0.6253</td>
<td>467.2790</td>
</tr>
<tr>
<td>Test set</td>
<td>8.2794</td>
<td>8.4440</td>
<td>8.1140</td>
<td>0.1638</td>
<td>41.3970</td>
</tr>
</tbody>
</table>

Table 4: Results of 3D-QSAR analysis using Random selection method (85%)

<table>
<thead>
<tr>
<th>TRIA L</th>
<th>TEST SET</th>
<th>kNN result</th>
<th>Simulated Annealing</th>
<th>Genetic Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14c,17a,4l,4q,4r,4w, 8b,8g,9</td>
<td>k Nearest Neighbour= 2 n = 49 Degree of freedom = 44 q2 = 0.6321 q2_se = 0.3913 Predr2 = 0.4170 pred_r2se = 0.3002</td>
<td>k Nearest Neighbour= 5 n = 49 Degree of freedom = 44 q2 = 0.1560 q2_se = 0.5926 Predr2 = 0.5958 pred_r2se = 0.3002</td>
<td>k Nearest Neighbour= 4 n = 49 Degree of freedom = 45 q2 = -0.1168 q2_se = 0.6817 Predr2 = -0.5758 pred_r2se = 0.5928</td>
</tr>
<tr>
<td>2</td>
<td>13a,13d,14a,19a,19d,4r,8a,8e,9</td>
<td>k Nearest Neighbour= 2 n = 49 Degree of freedom = 45 q2 = 0.4442 q2_se = 0.4529 Predr2 = -0.3759 pred_r2se = 0.8353</td>
<td>k Nearest Neighbour= 5 n = 49 Degree of freedom = 44 q2 = 0.0829 q2_se = 0.5818 Predr2 = 0.0176 pred_r2se = 0.7058</td>
<td>k Nearest Neighbour= 5 n = 49 Degree of freedom = 45 q2 = -0.1484 q2_se = 0.6510 Predr2 = 0.2088 pred_r2se = 0.6334</td>
</tr>
<tr>
<td>3</td>
<td>13d,14c,19d,4d,4j,4m,4p,4r,9</td>
<td>k Nearest Neighbour= 2 n = 49 Degree of freedom = 43 q2 = 0.7018 q2_se = 0.3404 Predr2 = -0.2799 pred_r2se = 0.7189</td>
<td>k Nearest Neighbour= 5 n = 49 Degree of freedom = 44 q2 = 0.1911 q2_se = 0.5606 Predr2 = 0.1789 pred_r2se = 0.7058</td>
<td>k Nearest Neighbour= 5 n = 49 Degree of freedom = 45 q2 = -0.0692 q2_se = 0.6446 Predr2 = 0.0974 pred_r2se = 0.6037</td>
</tr>
<tr>
<td>4</td>
<td>12,14d,18a,4c,4g,4i,4o,4t,9</td>
<td>k Nearest Neighbour= 4 n = 49 Degree of freedom = 44 q2 = 0.4627 q2_se = 0.4685 Predr2 = -0.6210 pred_r2se = 0.6606</td>
<td>k Nearest Neighbour= 5 n = 49 Degree of freedom = 44 q2 = 0.0950 q2_se = 0.6080 Predr2 = -0.8115 pred_r2se = 0.6983</td>
<td>k Nearest Neighbour= 5 n = 49 Degree of freedom = 45 q2 = -0.1504 q2_se = 0.6855 Predr2 = 0.4306 pred_r2se = 0.3915</td>
</tr>
<tr>
<td>5</td>
<td>15b,17a,18a,19d,4d,4q,8c,8e,9</td>
<td>k Nearest Neighbour= 4 n = 49 Degree of freedom = 44 q2 = 0.4752 q2_se = 0.4429 Predr2 = -0.6089 pred_r2se = 0.9035</td>
<td>k Nearest Neighbour= 5 n = 49 Degree of freedom = 44 q2 = 0.0775 q2_se = 0.5872 Predr2 = 0.3282 pred_r2se = 0.5839</td>
<td>k Nearest Neighbour= 5 n = 49 Degree of freedom = 45 q2 = -0.0979 q2_se = 0.6406 Predr2 = -0.0840 pred_r2se = 0.7416</td>
</tr>
</tbody>
</table>
Table 5: Results of 3D-QSAR analysis using Random selection method (80%)

<table>
<thead>
<tr>
<th>TRIAL</th>
<th>TEST SET</th>
<th>kNN result</th>
<th>Stepwise forward-backward</th>
<th>Simulated Annealing</th>
<th>Genetic Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| 1     | 12,14a, 14d, 17a, 19a, 19f, 4f, 4j, 4k, 4r, 8d, 9 | k Nearest Neighbour= 4  
n = 46  
Degree of freedom = 44  
q² = 0.2104  
q²_se = 0.5651  
Predr² = -0.6242  
pred_r²se = 0.7247 | k Nearest Neighbour= 5  
n = 46  
Degree of freedom = 41  
q² = 0.1169  
q²_se = 0.5977  
Predr² = -0.6818  
pred_r²se = 0.7374 | k Nearest Neighbour= 5  
n = 46  
Degree of freedom = 42  
q² = -0.1244  
q²_se = 0.6744  
Predr² = 0.2526  
pred_r²se = 0.4916 |
| 2     | 14c, 17a, 19e, 4g, 4q, 4r, 4w, 8b, 8e, 8g, 9 | k Nearest Neighbour= 4  
n = 46  
Degree of freedom = 44  
q² = 0.2400  
q²_se = 0.5714  
Predr² = -0.2694  
pred_r²se = 0.5290 | k Nearest Neighbour= 3  
n = 46  
Degree of freedom = 41  
q² = 0.0198  
q²_se = 0.6489  
Predr² = 0.3838  
pred_r²se = 0.3686 | k Nearest Neighbour= 4  
n = 46  
Degree of freedom = 42  
q² = -0.1119  
q²_se = 0.6911  
Predr² = -0.8464  
pred_r²se = 0.6380 |
| 3     | 12,13a, 13d, 14a, 19a, 19d, 4r, 4x, 8a, 8c, 8e, 9 | k Nearest Neighbour= 2  
n = 46  
Degree of freedom = 42  
q² = 0.4480  
q²_se = 0.4498  
Predr² = -0.0926  
pred_r²se = 0.7353 | k Nearest Neighbour= 3  
n = 46  
Degree of freedom = 41  
q² = 0.0990  
q²_se = 0.5746  
Predr² = -0.2460  
pred_r²se = 0.7852 | k Nearest Neighbour= 5  
n = 46  
Degree of freedom = 42  
q² = -0.1782  
q²_se = 0.6571  
Predr² = 0.2237  
pred_r²se = 0.6198 |
| 4     | 12,14d, 18a, 19c, 4c, 4i, 4o, 4r, 4s, 4t, 9 | k Nearest Neighbour= 5  
n = 46  
Degree of freedom = 40  
q² = 0.6547  
q²_se = 0.3739  
Predr² = -0.9512  
pred_r²se = 0.8038 | k Nearest Neighbour= 5  
n = 46  
Degree of freedom = 41  
q² = 0.0646  
q²_se = 0.6155  
Predr² = 0.2371  
pred_r²se = 0.5026 | k Nearest Neighbour= 5  
n = 46  
Degree of freedom = 42  
q² = -0.1603  
q²_se = 0.6855  
Predr² = 0.2633  
pred_r²se = 0.4939 |
| 5     | 10,14b, 17a, 18b, 4c, 4d, 4m, 4o, 4t, 4x, 8d, 9 | k Nearest Neighbour= 2  
n = 46  
Degree of freedom = 40  
q² = 0.5952  
q²_se = 0.4051  
Predr² = -0.9659  
pred_r²se = 0.8057 | k Nearest Neighbour= 5  
n = 46  
Degree of freedom = 41  
q² = 0.1104  
q²_se = 0.6006  
Predr² = -0.0129  
pred_r²se = 0.5783 | k Nearest Neighbour= 4  
n = 46  
Degree of freedom = 42  
q² = -0.1540  
q²_se = 0.6840  
Predr² = 0.1696  
pred_r²se = 0.5236 |

Figure 3: 3D-alignment of molecules.
Table 6: Results of 3D-QSAR analysis using Random selection method (75%)

<table>
<thead>
<tr>
<th>TRIAL</th>
<th>TEST SET</th>
<th>kNN result</th>
<th>Simulated Annealing</th>
<th>Genetic Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14c,17a,19e,4g,4i,4j,4l,4q,4r,4s,4w,8b,8e,8g,9</td>
<td>k Nearest Neighbor = 4 n = 43 Degree of freedom = 41 q2 = 0.2249 q2_se = 0.5844 Predr2 = -0.0006 pred_r2se = 0.4862</td>
<td>k Nearest Neighbor = 5 n = 43 Degree of freedom = 38 q2 = 0.0497 q2_se = 0.6471 Predr2 = -0.4015 pred_r2se = 0.5755</td>
<td>k Nearest Neighbor = 5 n = 43 Degree of freedom = 39 q2 = -0.1349 q2_se = 0.7072 Predr2 = -0.6110 pred_r2se = 0.6170</td>
</tr>
<tr>
<td>2</td>
<td>12,14d,17b,18a,18b,19c,4c,4g,4i,4o,4r,4s,4w,8e,9</td>
<td>k Nearest Neighbor = 2 n = 43 Degree of freedom = 37 q2 = 0.6428 q2_se = 0.3893 Predr2 = -0.9495 pred_r2se = 0.7431</td>
<td>k Nearest Neighbor = 2 n = 43 Degree of freedom = 38 q2 = 0.0871 q2_se = 0.6224 Predr2 = -0.7086 pred_r2se = 0.6957</td>
<td>k Nearest Neighbor = 5 n = 43 Degree of freedom = 39 q2 = -0.1350 q2_se = 0.6940 Predr2 = -0.1094 pred_r2se = 0.5605</td>
</tr>
<tr>
<td>3</td>
<td>10,14b,15a,17a,18b,4c,4d,4m,4o,4t,4x,8a,8d,8f,9</td>
<td>k Nearest Neighbor = 2 n = 43 Degree of freedom = 39 q2 = 0.4426 q2_se = 0.4760 Predr2 = 0.0867 pred_r2se = 0.5637</td>
<td>k Nearest Neighbor = 3 n = 43 Degree of freedom = 38 q2 = 0.1768 q2_se = 0.5784 Predr2 = 0.0557 pred_r2se = 0.5732</td>
<td>k Nearest Neighbor = 5 n = 43 Degree of freedom = 39 q2 = -0.2046 q2_se = 0.6997 Predr2 = 0.0663 pred_r2se = 0.5700</td>
</tr>
<tr>
<td>4</td>
<td>10,13d,14a,14b,14d,19c,20,4c,4d,4k,4q,4w,8b,8d</td>
<td>k Nearest Neighbor = 2 n = 43 Degree of freedom = 38 q2 = 0.6145 q2_se = 0.4046 Predr2 = -0.8133 pred_r2se = 0.7161</td>
<td>k Nearest Neighbor = 4 n = 43 Degree of freedom = 38 q2 = 0.1529 q2_se = 0.5998 Predr2 = -0.4238 pred_r2se = 0.6345</td>
<td>k Nearest Neighbor = 5 n = 43 Degree of freedom = 39 q2 = -0.1799 q2_se = 0.7079 Predr2 = 0.1903 pred_r2se = 0.4785</td>
</tr>
<tr>
<td>5</td>
<td>11,13c,15a,17a,18b,19a,19d,19e,4b,4d,4f,4j,4n,8c,9</td>
<td>k Nearest Neighbor = 2 n = 43 Degree of freedom = 40 q2 = 0.5015 q2_se = 0.4390 Predr2 = -0.5217 pred_r2se = 0.7820</td>
<td>k Nearest Neighbor = 5 n = 43 Degree of freedom = 38 q2 = 0.0915 q2_se = 0.5927 Predr2 = -0.1297 pred_r2se = 0.6738</td>
<td>k Nearest Neighbor = 5 n = 43 Degree of freedom = 39 q2 = -0.0324 q2_se = 0.6318 Predr2 = 0.3867 pred_r2se = 0.4965</td>
</tr>
</tbody>
</table>

CONCLUSION

Three dimensional quantitative structure activity relationship (3D QSAR) analysis using k nearest neighbor molecular field analysis (kNN MFA) method was performed on a series of arylbenzofuran derivatives as histamine H<sub>3</sub> receptor inhibitors using molecular design suite (VLifeMDS). This study was performed with 58 compounds (data set) using sphere exclusion (SE) algorithm and random selection method for the division of the data set into training and test set. kNN-MFA methodology with step wise variable selection forward-backward, Simulated Annealing and Genetic Algorithms methods were used for building the QSAR models. Two predictive models were generated with SW-kNN MFA. The most predictive model was generated by kNN (stepwise forward backward) using sphere exclusion data selection method. This model explains good internal (q<sup>2</sup> = 0.5445) as well as very good external (Predr<sup>2</sup> = 0.9013) predictive power of the model. The steric descriptors at the grid points, S<sub>2307</sub> and S<sub>2377</sub> plays important role in imparting activity. This model indicates that two steric descriptors are involved. The kNN-MFA contour plots provided further understanding of the relationship between structural features of substituted arylbenzofuran derivatives and their activities which should be applicable to design newer potential H<sub>3</sub> receptor inhibitors.

ACKNOWLEDGEMENT

The authors would like to thanks Head, SLT Institute of Pharmaceutical sciences, Guru Ghasidas University, Bilaspur (CG) for providing the necessary facilities.
REFERENCES

24. Liedtke S., Flau K., Kathmann M., Schlicker E., Stark H., Meier G., Schunack W., Replacement of imidazole by a piperidine moiety differentially affects the potency of