



RESEARCH PAPER

EXPLORING 2D AND 3D QSAR STUDIES OF INDOLE/ BENZOXIMIDAZOLE-5-CARBOXIMIDINE DERIVATIVES AS ANTICANCER AGENTS FOR THE DEVELOPMENT OF PREDICTIVE MODEL

Shivangi Agarwal¹, Vikash K. Mishra¹, Vivek Kumar², Mitali Mishra¹, Varsha Kashaw³ and Sushil K. Kashaw^{1*}

¹Department of Pharmaceutical Sciences, Dr. H.S. Gour Central University, Sagar-470 003, Madhya Pradesh, India

²Department of Pharmaceutical Chemistry, National Institute of Pharmaceutical Education and Research, Mohali, Chandigarh-160 062, India

³Department of Pharmaceutical Chemistry, Sagar Institute of Pharmaceutical Sciences, Sagar-470 228, Madhya Pradesh, India.

*E-mail: sushilkashaw@gmail.com
Tel.: +91 9425655720.

Received: Jul 09, 2015 / Revised: Oct 21, 2015 / Accepted: Oct 22, 2015

2D QSAR and 3D QSAR studies of indole/benzoximidazole-5-carboximidine derivatives as anti-cancer agents by Vlife Sciences MDS molecular modelling package has been performed to explore the physico-chemical properties and spatial disposition of the chemical groups towards the biological activity. The 2D-QSAR studies were carried out using the partial least squares (PLS) method coupled with stepwise variable selection, with $r^2 = 0.5960$ and $q^2 = 0.4062$ and the 3D-QSAR studies were performed using stepwise variable selection k-nearest-neighbour molecular field analysis (kNNMF) approach; with cross-validated correlation coefficient (q^2) of 0.6453 and a predicted r^2 for the external test ($pred_r^2$) of 0.7316.

Key words: QSAR, kNNMFA, PLS, Indole/benzoximidazole-5-carboximidine, Anticancer agents.

INTRODUCTION

Cancer is the second leading cause of death in the developed world. Cancer accounted 7.9 million deaths (around 13% of all deaths) in 2008 (Zali *et al* 2011). Regular use of some established screening tests can prevent the development of cancer through identification and removal or treatment of premalignant abnormalities (American Cancer Society, 2013). Currently, inspite of intensive research and some major advances in treatment, cancer claims the life of nearly one out of four. It is second to heart diseases responsible for 35% of deaths in United States (Copper, 1992). The failure of the curative

treatment of cancer patients often occurs as a result of intrinsic or acquired drug resistance of the tumor to chemotherapeutic agents. The resistance of tumors occurs not only to a single cytotoxic drug used, but also occurs as a cross-resistance to a whole range of drugs with different structures and cellular targets (Ozben, 2006). There are more than 100 distinct types of cancer and subtypes of tumors can be found within specific organs (Hanahan and Weinberg, 2000). The urokinase (uPA) pathway of plasminogen activation plays a key role in tissue remodelling, cancer cell invasion and in metastasis. uPA is one of the serine proteases

that catalyzes the conversion of plasminogen to plasmin, an active enzyme that is able to degrade a variety of extracellular matrix proteins. The involvement of urokinase-type plasminogen activator receptor in the pathology of human cancers is well documented. High levels of uPAR in tumour tissues and plasma from patients with various human cancers are associated with poor prognosis and increased risk of tumour recurrence and metastasis. The uPAS (urokinase plasminogen activating system) is involved in the extracellular conversion of the ubiquitous inactive plasminogen to the broad-spectrum serine protease plasmin, implicated in numerous pathophysiological process requiring the remodelling of extracellular matrix (ECM) and basement membranes (BM) (Ulisse *et al* 2009). Inhibition of expression of these components leads to a reduction in the invasive and metastatic capacity of many tumors (Choong and Nadesapillai, 2003).

Quantitative structure-activity relationship models (QSAR models) are regression or classification models used in the chemical and biological sciences and engineering. Like other regression models, QSAR regression models establish a linear relationship between a set of molecular descriptors of chemicals and the

biological activity of the chemical while classification QSAR models shows a non-linear relationship between a set of molecular descriptors of chemicals and the biological activity of the chemical (Bansal *et al* 2011; Kumar, 2011; Sharma *et al* 2011; Jain *et al* 2013; Pahwa and Papreja, 2014).

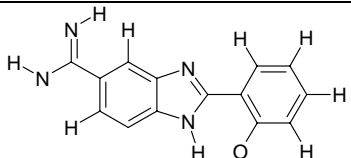
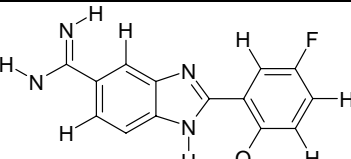
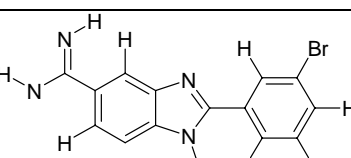
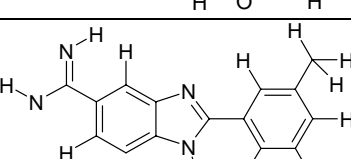
QSAR studies can reduce the costly failures of drug candidates in clinical trials by filtering the combinatorial libraries (Sethi, 2012). In continuation of the research of molecular modelling, in this study, we report the 2D and 3D QSAR analysis of indole/benzoximidazole-5-carboximidine derivatives as anti-cancer agents by VlifeSciences MDS molecular modeling package.

MATERIALS AND METHODS

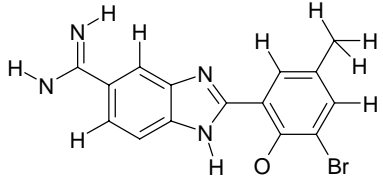
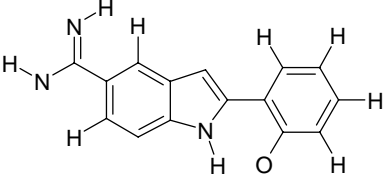
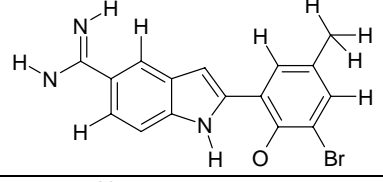
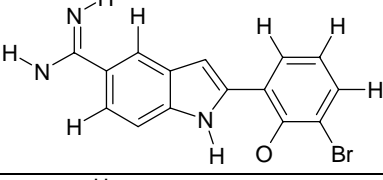
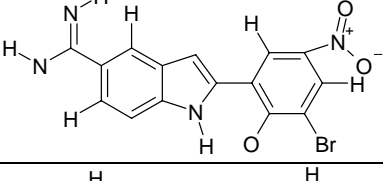
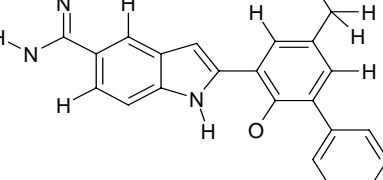
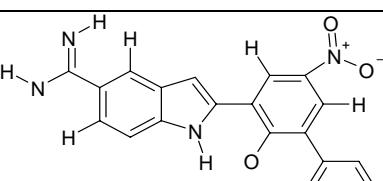
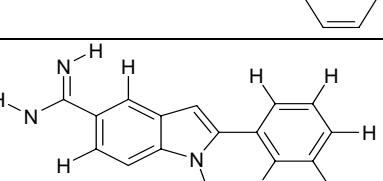
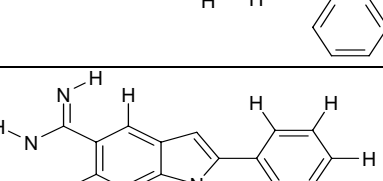
Selection of molecules and Data set

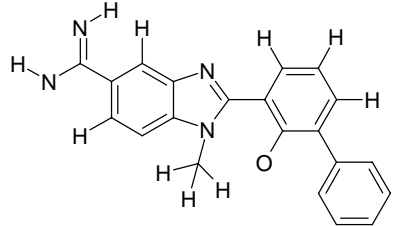
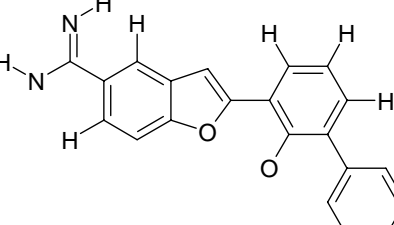
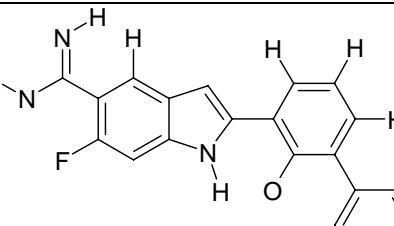
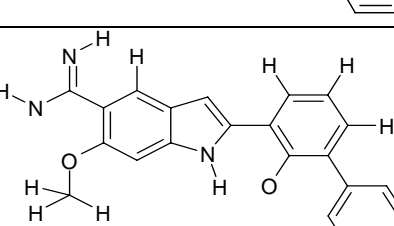
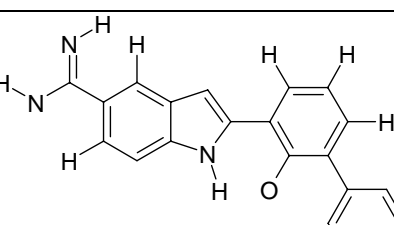
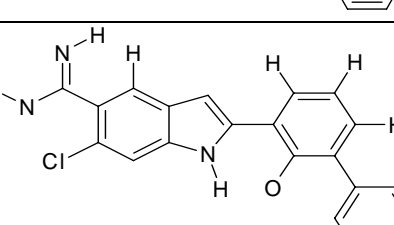
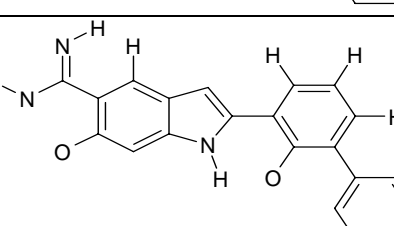
A dataset of thirty nine indole/benzoximidazole-5-carboximidine derivatives (Bhongade and Gadad, 2004) reported to have inhibitory activity on urokinase plasminogen activator (uPA) was used in the present 2D and 3D QSAR analysis. The inhibitory concentration values reported as pIC_{50} ($-\log IC_{50}$) were used for the studies. The structures along with their pIC_{50} values are represented in **Table 1**.

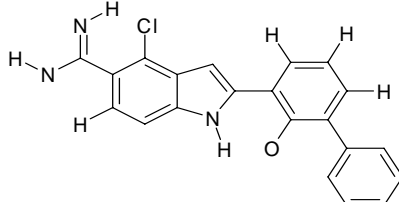
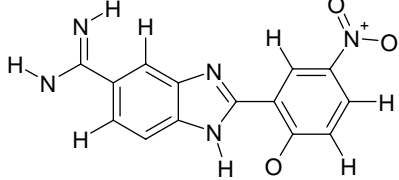
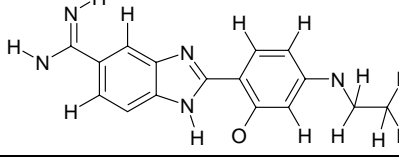
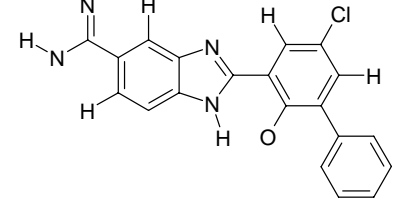
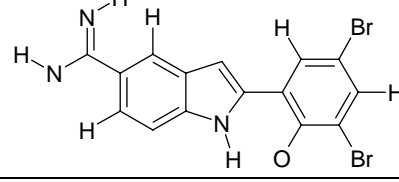
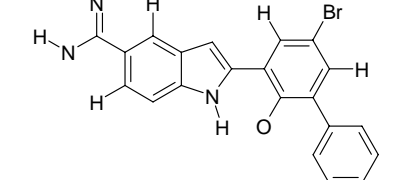
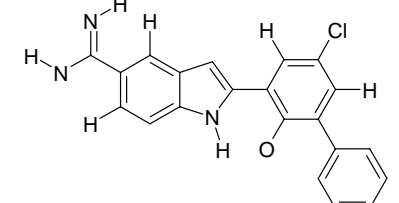
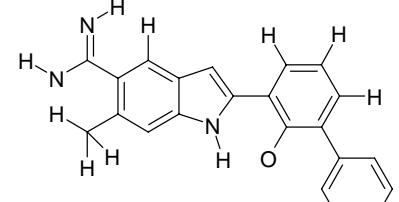
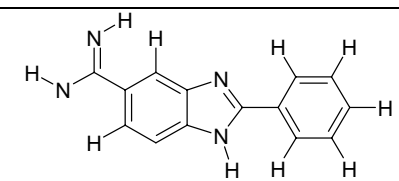
Table 1. Chemical structures of data set used for 2D and 3D QSAR Studies with actual and predicted activity from the best model

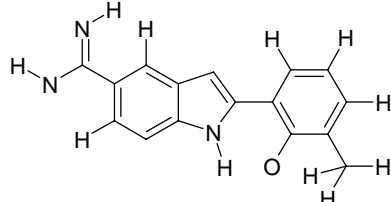
| Molecule | Structures | Activity(pIC_{50}) | |
|----------|---|------------------------|-----------|
| | | Actual | Predicted |
| S1 |  | -0.74 | -0.81 |
| S2 |  | -0.447 | -0.55 |
| S3 |  | -0.568 | -0.64 |
| S4 |  | -0.87 | -0.44 |

| | | | |
|-----|--|--------|-------|
| S5 | | 0.556 | 0.89 |
| S6 | | 0.045 | -0.32 |
| S7 | | 0.259 | -0.09 |
| S8 | | 0.25 | 0.36 |
| S9 | | -0.556 | 0.02 |
| S10 | | 0.39 | 0.92 |
| S11 | | -0.556 | -0.80 |
| S12 | | -0.278 | -0.05 |
| S13 | | -0.591 | -0.48 |

| | | | |
|-----|---|--------|-------|
| S14 |  | 0.552 | 0.52 |
| S15 |  | -0.38 | -0.31 |
| S16 |  | 1.25 | 0.79 |
| S17 |  | 1.45 | 0.51 |
| S18 |  | 0.552 | 0.84 |
| S19 |  | 1.42 | 1.70 |
| S20 |  | 1.602 | 1.73 |
| S21 |  | -1.113 | -1.02 |
| S22 |  | -0.5 | -0.82 |

| | | | |
|-----|---|--------|-------|
| S23 |  | -1.041 | -0.80 |
| S24 |  | 0.259 | 0.25 |
| S25 |  | 1.698 | 1.75 |
| S26 |  | 0.44 | 0.24 |
| S27 |  | 2.096 | 1.29 |
| S28 |  | 2.045 | 1.98 |
| S29 |  | -0.3 | -0.21 |

| | | | |
|-----|---|--------|-------|
| S30 |  | -0.477 | 0.70 |
| S31 |  | -0.518 | -0.48 |
| S32 |  | -1.13 | -1.25 |
| S33 |  | 0.301 | 0.61 |
| S34 |  | 1 | 0.84 |
| S35 |  | 1.36 | 1.59 |
| S36 |  | 1.25 | 1.54 |
| S37 |  | 1 | 1.09 |
| S38 |  | -0.903 | -1.06 |

| | | | |
|-----|---|-------|-------|
| S39 |  | -1.74 | -0.36 |
|-----|---|-------|-------|

2D QSAR

All the compounds for 2D QSAR was subjected to energy minimization by MMFF (Molecular mechanics force field method). Various 2D descriptors like topological, physicochemical, alignment-independent descriptors were calculated after which by invariable column was removed and the training and test set was selected by Manual Selection Method. The model for the 2D-QSAR study was generated using PLS with forward backward as the variable selection method.

3D QSAR

3D-QSAR refers to the application of force field calculations requiring three-dimensional structures based on protein crystallography or molecule superimposition (Ibezim *et al* 2009). For studies, the molecules are converted from 2D to 3D structures, optimized by MMFF (Molecular mechanics force field method) and then were aligned using Template based Alignment method by taking most active molecule (**Figure 1**) as the reference molecule and basic moiety (**Figure 2**) as the template. The alignment is shown in **Figure 3**. The QSAR model was built by KNN method using forward-backward as variable selection method. It examines the steric fields and the electrostatic fields.

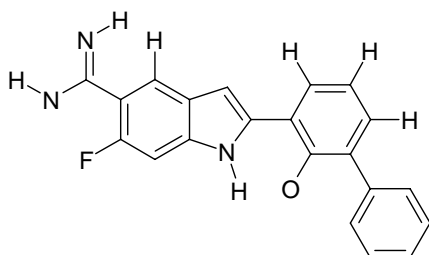


Fig. 1. Reference molecule used for alignment by template based alignment

RESULTS AND DISCUSSION

2D QSAR

The best QSAR model was selected among the various models generated by PLS (Partial Least Square analysis). The 2D QSAR equation and the statistical parameters generated are shown in

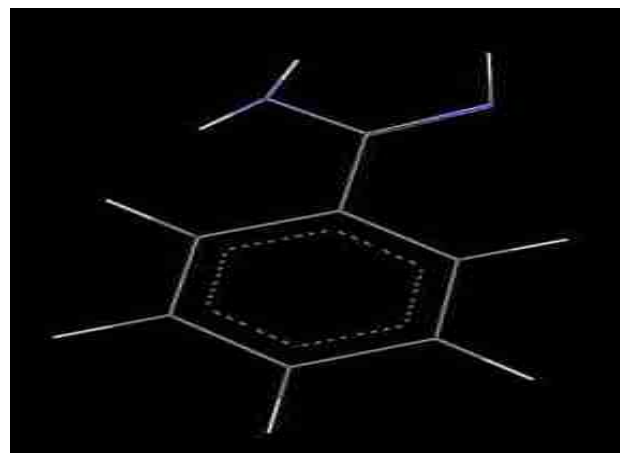


Fig. 2. Basic moiety as a template for alignment

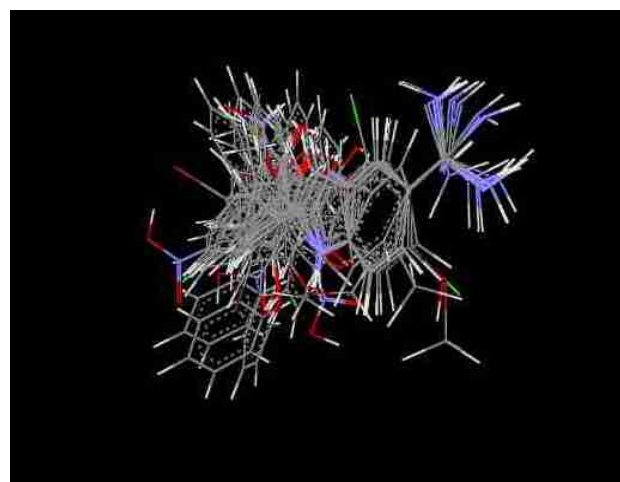


Fig. 3. 3D view of aligned molecules

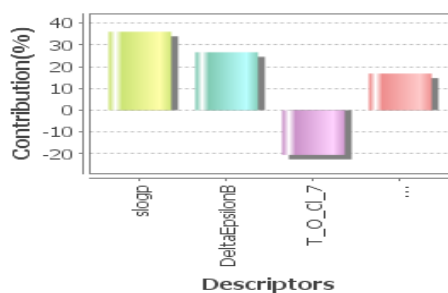
Table 2. The QSAR model showed $r^2 = 0.5960$ and standard error of 0.6907. The stability of model judged by leave-one-out procedure ($q^2 = 0.4062$) suggesting that the models can be useful for meaningful predictions. Furthermore, the predictive potential of model was observed by the $\text{pred}_r^2 = 0.1208$. The descriptors contributing to the biological activity as generated were slogP, EpsilonB, T_O_Cl_7 and H-Donor. The correlation matrix given by **Table 3** and the values less than 0.6 indicated the absence of multi colinearities in the model. **Figure 4** gives a pictorial representation of different 2D parameters and their contributions towards anticancer activity.

Table 2. Statistical results of 2D QSAR equation generated by PLS method

| | | | |
|-------------------|--|---------------------------|---------------------------------|
| Equation | $pIC_{50} = 1.05 \text{ slogP} + 22.7091 \text{ Delta EpsilonB} - 1.9378 \text{ T_O_Cl_7} + 1.1410 \text{ H-Donor count} - 10.1773$ | | |
| Statistics | n = 33 | Degree of freedom = 29 | F test = 14.2619 |
| | $r^2 = 0.5960$ | $q^2 = 0.4062$ | pred_r ² = 0.1208 |
| | $r^2 \text{ se} = 0.6907$ | $q^2 \text{ se} = 0.8373$ | pred_r ² se = 0.7331 |

Table 3. Correlation matrix for the descriptors contributing to the 2D QSAR model

| | | | |
|--------------------|-----------------|--------------------|--------------|
| | Mol. wt. | SdsNE-index | Score |
| Mol. wt. | 1 | -0.413695 | 1 |
| SdsNE-index | -0.413695 | 1 | 1 |

**Fig. 4.** Contribution chart of selected 2D descriptors for anticancer activity**slogp**

This descriptor signifies log of the octanol/water partition coefficient

T_O_Cl_5

This is the count of number of Oxygen atoms (single double or triple bonded) separated from chlorine atom by 5 bond distance in a molecule.

H-Donor count

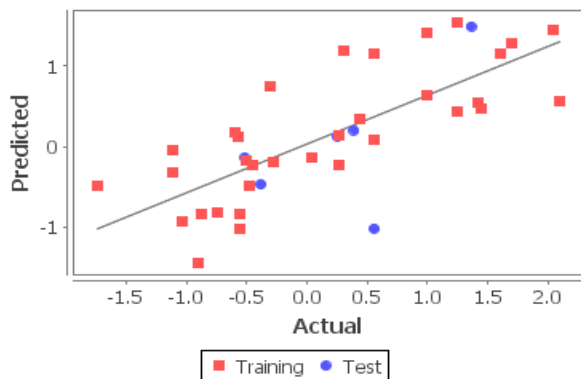
Number of hydrogen bond donor atoms

Uni-column statistics as shown in **Table 4** for training and test set, were generated to check correctness of selection criteria for training and test set molecules. Higher standard deviation in training set indicates wide distribution of activity of molecules as compared to test set molecules.

Figure 5 shows the graph plotted between actual and predicted biological activity. Training set (points in red) as well as test set (points in blue) are well closed to regression line and also the training set are encircling the test set showing good predictive ability of the model. The predicted activity values for the compounds in the training set and test set, along with their corresponding observed activity values, are given in **Table 5**.

Table 4. Uni-column statistics for training set and test set

| Set | Column Name | Average | Max | Min | Std Dev | Sum |
|----------|-------------|---------|--------|---------|---------|--------|
| Training | pIC_{50} | 0.1645 | 2.0960 | -1.7400 | 1.0345 | 5.4290 |
| Test | pIC_{50} | 0.2763 | 1.3600 | -0.5180 | 0.6823 | 1.6580 |

**Fig. 5.** Fitness plot between observed and predicted biological activities of training (blue spot) and test (red spot) molecules of 2D model.**3D QSAR**

3D-QSAR was used to optimize the steric requirement. 3D data points were generated that contributed to k Nearest Neighbour Molecular Field Analysis (kNN MFA) 3D-QSAR model. This is followed by generation of a common rectangular grid around the molecules. The steric interaction energies were computed at the lattice points of the grid using a methyl probe of charge +1. The data points generated by 3D-QSAR are shown in **Figure 6**. The best model generated with kNN MFA method showed a q^2 , pred_r² and K nearest neighbour as 0.8105, 0.7015 and 2 respectively as shown in **Table 6**. The predicted activity values for the compounds

Table 5. Observed and predicted activity of training set and test set with their residuals of 2D QSAR

| Molecule | Actual BA | Predicted BA | Residual |
|----------|-----------|--------------|----------|
| S1 | -0.74 | -0.813449 | 0.073449 |
| S2 | -0.447 | -0.219817 | -0.22718 |
| S3 | -0.568 | 0.130168 | -0.69817 |
| S4 | -0.87 | -0.841612 | -0.02839 |
| S5* | 0.556 | -1.01552 | 1.57152 |
| S6 | 0.045 | -0.131271 | 0.176271 |
| S7 | 0.259 | -0.219817 | 0.478817 |
| S8* | 0.25 | 0.130168 | 0.119832 |
| S9 | -0.556 | -1.01552 | 0.45952 |
| S10* | 0.39 | 0.206531 | 0.183469 |
| S11 | -0.556 | -0.841612 | 0.285612 |
| S12 | -0.278 | 0.183642 | -0.46164 |
| S13 | -0.591 | 0.080751 | -0.67175 |
| S14 | 0.552 | 0.080751 | 0.471249 |
| S15* | -0.38 | -0.466336 | 0.086336 |
| S16 | 1.25 | 0.446082 | 0.803918 |
| S17 | 1.45 | 0.477281 | 0.972719 |
| S18 | 0.552 | 1.15872 | -0.60672 |
| S19 | 1.42 | 0.55255 | 0.86745 |
| S20 | 1.602 | 1.15587 | 0.44613 |
| S21 | -1.113 | -0.037207 | -1.07579 |
| S22 | -0.5 | -0.170716 | -0.32928 |
| S23 | -1.041 | -0.922126 | -0.11887 |
| S24 | 0.259 | 0.143706 | 0.115294 |
| S25 | 1.698 | 1.29298 | 0.40502 |
| S26 | 0.44 | 0.337396 | 0.102604 |
| S27 | 2.096 | 0.561509 | 1.534491 |
| S28 | 2.045 | 1.45917 | 0.58583 |
| S29 | -3 | 0.752633 | -3.75263 |
| S30 | -0.477 | -0.478671 | 0.001671 |
| S31* | -0.518 | -0.131271 | -0.38673 |
| S32 | -1.113 | -0.314839 | -0.79816 |
| S33 | 0.301 | 1.196050 | -0.89505 |
| S34 | 1 | 1.420900 | -0.4209 |
| S35* | 1.36 | 1.495950 | -0.13595 |
| S36 | 1.25 | 1.551030 | -0.30103 |
| S37 | 1 | 0.638880 | 0.36112 |
| S38 | -0.903 | -1.43510 | 0.5321 |
| S39 | -1.74 | -0.476281 | -1.26372 |

*Test set

in the training set and test set, along with their corresponding observed activity values, are given in **Table 7** and the fitness plot in **Figure 7** which shows the graph plotted between actual and predicted biological activity. Training set (points in red) as well as test set (points in blue) are well closed to regression line and showing

good predictive ability of the model. The ranges of data point values were based on the variation of the field values at the chosen points using the most active molecule and its nearest neighbour set. Points generated in kNN MFA 3D-QSAR model were S_1012 (30.0000, 30.0000), E_547 (-0.2724, 0.3311) *i.e.* steric and electrostatic data

Table 6. Statistical results of 3D QSAR model generated by SW kNN MFA method

| Model summary | | | | | | | |
|-------------------------------------|--------------------------|--------|------------------------|-------------------------|-----------------------------|------------------------------|---------------------------------|
| kNN Method | Training Set Size = 31 | | | | Test Set Size = 8 | | |
| Statistics | k Nearest Neighbour= 2 | n = 31 | Degree of freedom = 28 | q ² = 0.6453 | q ² _se = 0.5914 | Pred_r ² = 0.7316 | pred_r ² se = 0.5230 |
| Selected descriptor and their range | S_1012 (30.0000 30.0000) | | | | E_547 (-0.2724 0.3311) | | |

Table 7. Observed and predicted activity of training set and test set with their residuals of 3D QSAR

| Molecule | Actual BA | Predicted BA | Residual |
|----------|-----------|--------------|----------|
| S1 | -0.74 | -0.819529 | 0.073449 |
| S2 | -0.447 | -0.377798 | -0.22718 |
| S3 | -0.568 | 0.306998 | -0.69817 |
| S4* | -0.87 | -0.647808 | -0.02839 |
| S5* | 0.556 | 0.439421 | 1.57152 |
| S6 | 0.045 | 0.057331 | 0.176271 |
| S7 | 0.259 | 1.12333 | 0.478817 |
| S8 | 0.25 | 0.302605 | 0.119832 |
| S9* | -0.556 | -0.158517 | 0.45952 |
| S10 | 0.39 | -0.647088 | 0.183469 |
| S11 | -0.556 | -0.916909 | 0.285612 |
| S12 | -0.278 | -0.462462 | -0.46164 |
| S13 | -0.591 | 0.335349 | -0.67175 |
| S14 | 0.552 | -0.168048 | 0.471249 |
| S15* | -0.38 | -0.527896 | 0.086336 |
| S16 | 1.25 | 0.636405 | 0.803918 |
| S17 | 1.45 | 1.00428 | 0.972719 |
| S18 | 0.552 | 0.857929 | -0.60672 |
| S19 | 1.42 | 0.478506 | 0.86745 |
| S20 | 1.602 | 1.16161 | 0.44613 |
| S21 | -1.113 | -1.34338 | -1.07579 |
| S22 | -0.5 | -0.20481 | -0.32928 |
| S23 | -1.041 | -1.42854 | -0.11887 |
| S24 | 0.259 | 0.957433 | 0.115294 |
| S25 | 1.698 | 0.897315 | 0.40502 |
| S26* | 0.44 | -0.373997 | 0.102604 |
| S27 | 2.096 | 1.12273 | 1.534491 |
| S28* | 2.045 | 1.00109 | 0.58583 |
| S29 | -3 | -0.462813 | -3.75263 |
| S30 | -0.477 | -0.360151 | 0.001671 |
| S31* | -0.518 | -0.664076 | -0.38673 |
| S32 | -1.113 | -0.647785 | -0.79816 |
| S33 | 0.301 | -0.736929 | -0.89505 |
| S34 | 1 | 0.757103 | -0.4209 |
| S35* | 1.36 | 1.12951 | -0.13595 |
| S36 | 1.25 | 1.14399 | -0.30103 |
| S37 | 1 | 1.30164 | 0.36112 |
| S38 | -0.903 | -0.131259 | 0.5321 |
| S39 | -1.74 | -0.908508 | -1.26372 |

*Test set

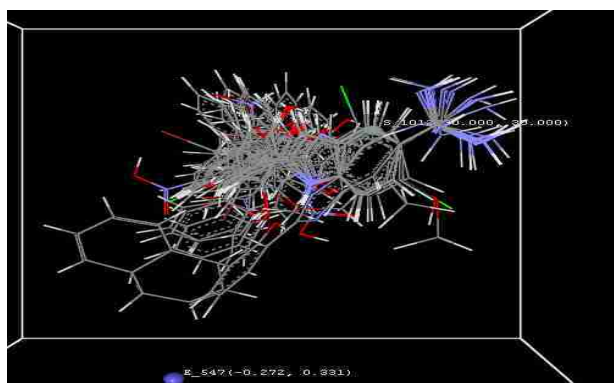


Fig. 6. Relative positions of the steric and electrostatic fields around aligned molecules

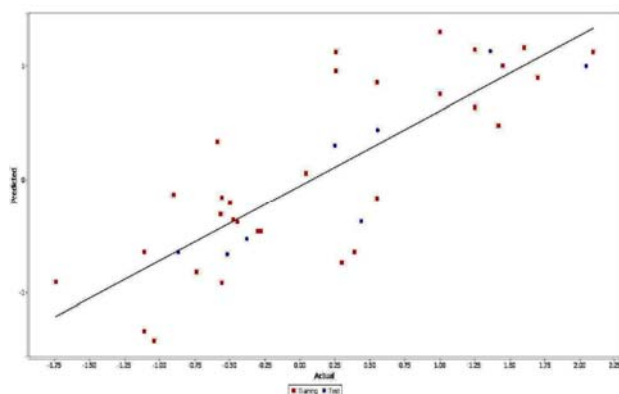


Fig. 7. Fitness plot between observed and predicted activities of the training (blue spot) and test (red spot) molecules of 3D model

points at lattice points 1012 and 547 respectively. Positive values in steric field descriptors indicated the requirement of positive steric potential for enhancing the biological activity of indole/benzoximidazole-5-carboximidine derivatives. Negative value in electrostatic potential indicated the requirement of negative electrostatic potential for enhancing the biological activity. The positive value of

REFERENCES

- American Cancer Society. Cancer Prevention and Early Detection Facts & Figures, 2013.
- Bansal H, Sharma A, Sharma V, Kumar V. Pharmacophore modeling studies on xanthenes as monoamine oxidase-A inhibitors. *Bull. Pharm. Res.* 2011;1(1):15-21.
- Bhongade BA, Gadad AK. 3D-QSAR CoMFA/CoMSIA studies on urokinase plasminogen activator (uPA) inhibitors: a strategic design in novel anticancer agents. *Bioorg. Med. Chem.* 2004;12(10):2797-805. [DOI: 10.1016/j.bmc.2004.02.019]
- Choong PF, Nadesapillai AP. Urokinase plasminogen activator system: a multifunctional role in tumor progression and metastasis. *Clin. Orthop. Relat. Res.* 2003;415(Suppl):S46-58.
- Copper GM. Elements of Human Cancer. Jones and Barlett Learning, Burlington, Massachusetts, United States: 1992.

steric factors showed that more steric substituents were preferred at the position of generated data points S_1012 (30.0000, 30.0000). Similarly the negative values of electrostatic descriptor suggested the requirement of more less electrostatic substituents at position of generated data point E_547 (-0.2724, 0.3311) for maximum activity.

CONCLUSION

2D and 3D studies were performed on indole/benzoximidazole-5-carboximidine derivatives for their anticancer activity. 2D QSAR study indicated the requirement of slogP, delta epsilonB and H-Donor count which positively contributed to the biological activity and removal of T_O_Cl_7 which negatively contributed to the biological activity. Points generated in SA kNN-MFA 3D-QSAR model are S_1012 (30.0000, 30.0000), E_547 (-0.2724, 0.3311) *i.e.* steric and electrostatic data points at lattice points 1012 and 547 respectively. 3D-QSAR gave information about nature of substituents like more steric substituents are required at data points S_1012 (30.0000, 30.0000) and less electrostatic substituents are required at data point S_1161 (30.0000, 30.0000) for maximum activity. Above studies rendered significant information which gives detailed structural insights that can provide crucial clues and guidance that can be used in the successful designing of novel highly active analogues against uPA.

ACKNOWLEDGEMENT

One of the author wish to acknowledge to the AICTE (All India Council of Technical Education) for granting of the fund for the completion of this project work.

- Hanahan D, Weinberg RA. The hallmarks of cancer review. *Cell* 2000; 100(1): 57-70.
- Ibezim EC, Duchowicz PR, Ibezim EN, Mullen LMA, Onyishi IV, Brown SA, Castro EA. Computer-aided linear modeling employing QSAR for drug discovery. *Afr. J. Basic Appl. Sci.* 2009;1(3-4):76-82.
- Jain J, Bansal SK, Chowdhury P, Sinha R, Tripathi U, Malhotra M. *In silico* pharmacophore validation of anticonvulsant activity of (*E*) (\pm)-3-menthone derivatives. *Bull. Pharm. Res.* 2013;3(3):146-56.
- Kumar V. Topological models for the prediction of tyrosine kinase inhibitory activity of 4-anilinoquinazolines. *Bull. Pharm. Res.* 2011;1(2):53-9.
- Ozben T. Mechanisms and strategies to overcome multiple drug resistance in cancer. *FEBS Lett.* 2006;580(12):2903-9. [DOI: 10.1016/j.febslet.2006.02.020]

Pahwa P, Papreja M. Validation of SOMFA using data mining technique. *Int. J. Soft Comput. Engg.* 2014;4:1-4.
Sethi NS. A review on computational methods in developing quantitative structure-activity relationship (QSAR). *Int. J. Drug Res. Tech.* 2012;2(4S):313-41.
Sharma V, Wakode SR, Lather V, Mathur R, Fernandes MX. Structure based rational drug design of selective phosphodiesterase-4 ligands as anti-inflammatory

molecules. *Bull. Pharm. Res.* 2011;1(2):33-40.
Ulisse S, Baldini E, Sorrenti S, D'Armiento M. The urokinase plasminogen activator system: A target for anti-cancer therapy. *Curr. Cancer Drug Targets* 2009;9(1):32-71. [DOI: 10.2174/156800909787314002]
Zali H, Rezaei-Tavirani M, Azodi M. Gastric cancer: prevention, risk factors and treatment. *Gastroenterol. Hepatol. Bed Bench.* 2011;4(4):175-85.
