ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 8, Issue, 10, pp.39618-39623, October, 2016

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

COMPARATIVE QSAR ANALYSIS OF HISTONE DEACETYLASE 6 (HDAC6) INHIBITORS AS ANTI-CANCER AGENTS

***Sandhya Vijayasarathy and Jhinuk Chatterjee**

Department of Biotechnology, PES Institute of Technology, Bangalore, India

ARTICLE INFO

Article History:
Received 20th July, 2016
Received in revised form
05th August, 2016
Accepted 28th September, 2016
Published online 30th October, 2016

Key words:

QSAR,
Histone Deacetylase,
Anti-cancer,
Regression,
Molecular Field Analysis.

ABSTRACT

Histone deacetylases (HDACs) remain a promising class of anti-cancer drug targets with an ability to reverse abnormal epigenetic states associated with cancer. HDAC6, a subtype of HDAC, functions at the crossroads between atleast two cell signalling pathways involving ubiquitination and lysine acetylation. Over expression of this enzyme is associated with tumorigenesis and cell survival, other than promoting metastasis in cancer cells. In this study, a comparative quantitative structure activity relationship (QSAR) analyses has been performed on HDAC6 inhibitors for predicting their inhibitory activity using two-dimensional and three-dimensional QSAR models. 2D QSAR models were built using Multiple Linear Regression (MLR), Principal Component Regression (PCR) besides Partial Least Squares regression (PLS) methods, in addition to a 3D QSAR model which was developed using k-Nearest Neighbor Molecular Field Analysis (kNN-MFA). Among all the developed models, multiple linear regression (MLR) model performed better with the correlation coefficient $r^2 = 0.7381$ and cross-validated squared correlation coefficient $q^2 = 0.6449$ with external predictive ability of pred_$r^2 = 0.5107$. Thus, the information rendered by these QSAR models may lead to a better understanding of structural requirements of this class of compounds against cancer in addition to paving the way for design of new and potent histone deacetylase inhibitors.

Citation: Sandhya Vijayasarathy and Jhinuk Chatterjee, 2016. "Comparative QSAR analysis of Histone Deacetylase 6 (HDAC6) inhibitors as Anti-cancer agents", *International Journal of Current Research*, 8, (10), 39618-39623.

INTRODUCTION

It is well established that Histone deacetylase (HDAC) enzyme is involved in the deacetylation of α-acetyl lysine that resides within the NH_2 terminal tail of histones resulting in inhibition of gene transcription (Dokmanovic *et al*., 2013). HDACs are found over-expressed in different types of cancer, and is considered as a major target for epigenetic therapy. Various studies have shown that HDAC inhibition elicits anti-cancer effects in many tumor cells by inhibiting cell growth and initiating differentiation of cells. The development of HDAC inhibitors as anti-cancer drugs has begun and compounds like Trichostatin A (TSA), Suberanilohydroxami acid (SAHA), Apicidin, Trapoxin along with synthetic inhibitors have been studied. Cell-based studies have revealed that HDAC inhibitors have a dominant anti-proliferative property, triggering cell-cycle arrest, apoptosis, and differentiation. These anti-proliferative effects were more distinct in tumor cells when compared to normal cells. HDACs have become chief targets and the quest for HDAC inhibitors has intensified, gathering excessive attention in drug discovery over the years (Sharma and Sharma, 2015). HDAC6, a class IIB member of HDAC, disturbs transcription and translation by regulating Heat Shock Protein 90 (HSP90) and stress granules respectively. Over-expression of this enzyme is associated with tumorigenesis and cell survival, in addition to promoting metastasis in cancer cells (Masangkay and Sakamoto, 2011). Selective inhibition of HDAC6 can be one of the promising approaches for the treatment of cancer.

Quantitative Structure Activity Relationship (QSAR) modelling is vastly employed in medicinal chemistry (Cherkasov *et al*., 2014). QSARs are mathematical models that establish relationship between molecular structures of compounds and their biological activities in a quantitative manner (Wong *et al*., 2014). QSAR works on the assumption that structurally alike compounds possess similar activities. These models can be developed by using various supervised and/or unsupervised machine learning techniques (Vijayasarathy and Chatterjee 2015, Ventura *et al*., 2013). QSAR analysis enables researchers to pick out the most promising compounds from a large compound library before subjecting them to biological testing, thus minimizing the number of time-consuming, expensive and laborious

Corresponding author: Sandhya Vijayasarathy,
Department of Biotechnology, PES Institute of Technology, Bangalore, India.

experiments (Wong *et al*., 2014). Some of the applications of QSAR models include: prediction of the biological activity (e.g., IC_{50}), classify compounds into classes (e.g., Inhibitor versus Non-inhibitors), analysis of structural characteristics that can give rise to the properties of interest, modelling of ADME parameters, lead compound optimization and diagnosis of mechanism of drug action, thus becoming beneficial for screening promising compounds having robust properties. QSAR provides drug designers with information that might be of use to improve the efficacy of drugs. k-nearest neighbour, 3D methods like Comparative molecular field analysis (COMFA) and comparative molecular similarity analysis (COMSIA) continue to be some of the QSAR techniques that are being incorporated in modern QSAR research. QSAR techniques are largely used in medical research to correlate the molecular structures with activity predictions, thereby screening compounds with undesirable properties or features (Sharma, 2015). In this study, QSAR models were developed using 2D and 3D-QSAR approaches in addition to finding the structural or chemical features required for HDAC6 inhibitory activity which can be further used for designing more potent anti-cancer compounds.

MATERIALS AND METHODS

Computational Details

All the computational studies were performed using VLife MDS 4.4 (Molecular Design Suite) software provided by V-life Sciences Technologies Pvt. Ltd., Pune, India (http://www.vlifesciences.com/).

Dataset

A data set comprising of forty histone deacetylase 6 inhibitors was used for the present QSAR study. These inhibitors with known inhibitory activity (IC_{50}) value were collected from PubChem Bioassay database (Wang *et al*., 2012) and their structures were subsequently downloaded as 3D SDF (Structure-Data File). Lipinski's rule of five was employed and no violation was found, which means that the compounds possess good pharmacokinetic profile. The inhibitory activity values were converted to negative logarithmic scale (pIC_{50}) and then used as the dependent variable for the QSAR analysis.

Computation of 2D descriptors

The Physiochemical as well as Alignment independent (AI) descriptors were calculated using V-Life MDS 4.4. For calculation of AI descriptors, every atom in the molecule was assigned at least one and at most three attributes. AI descriptors were computed using the following attributes: 2 (double bonded atom), 3 (triple bonded atom), C, N, O, S, H, F, Cl, Br and I with the distance range of 0–7. A total of 680 descriptors were calculated. The pre-processing of the 2D descriptors was done by removing the invariables (constant column) along with highly correlated ones, resulting in 350 descriptors.

Selection of training and test set

A dataset of forty molecules was divided randomly into training set and test set for generating 2D QSAR models for predicting the inhibitory activity of HDAC6 inhibitors. Selection of the training set and test set molecules was done based on the structural diversity and a wide range of activity was included. The maximum and minimum values in training and test set were compared in a way that: The maximum value of pIC_{50} of test set must be less than or equal to maximum value of pIC_{50} of training set and the minimum value of pIC_{50} of test set must be higher than or equal to minimum value of pIC_{50} of training set.

Calculation of 3D-QSAR descriptors

The molecules were aligned by template-based method (Ajmani *et al*., 2006, Sarankar and Pathak, 2012). The template structure, i.e., hydroxamic acid moiety was used for alignment by considering the common elements of the series. The compound with high activity, which makes a valid lead molecule, was selected as the reference molecule. Both steric and electrostatic fields were calculated at each lattice point of a regularly spaced grid box of 2 Å. A methyl probe of charge +1.0 with 10.0 kcal/mole electrostatic and 30.0 kcal/mole steric was used for generating the fields. This resulted in 5000 field descriptors (2500 for each field type).

Variable selection and construction of QSAR models

2D QSAR models were developed using Multiple Linear Regression (MLR), Principal Component Regression (PCR) and Partial Least Square (PLS) techniques with stepwise forward-backward selection method. 3D QSAR models were developed using kNN-MFA with stepwise forward-backward selection method.

Model validation

Leave-one-out (LOO) Cross Validation was used as internal validation method. In this method, each molecule in the training set was excluded once and the activity of the eliminated molecule was predicted by using the model developed by the remaining molecules. The value of q^2 was calculated using Equation 1 given below, which describes the internal stability of the model.

$$q^2 = 1-\sum(y_i - \hat{y}_i)^2 / \sum(y_i - y_{mean})^2 \qquad \text{.....(1)}$$

Where, yi and $\hat{y}_i$ = actual and predicted activity of the ith molecule in the training set respectively, and y_{mean} = average activity of training set molecules (Sharma and Sharma, 2015). For external validation, activity of each test set molecule was predicted using the model built on the training set. The pred_r^2 value was calculated using Equation 2.

$$pred\_r^2 = 1-\sum(y_i - \hat{y}_i)^2 / \sum(y_i - y_{mean})^2 \qquad \text{.....(2)}$$

Where, yi, $\hat{y}_i$ = actual and predicted activity of the ith molecule in the test set respectively, and y_{mean} = average activity of training set molecules.

RESULTS AND DISCUSSION

Quantitative structure activity relationship plays an imperative role in unearthing of new and potent chemical entities. In this study, 2D QSAR models were developed for predicting inhibitory activity of various HDAC6 inhibitors using MLR, PCR and PLS techniques along with 3D QSAR

modelling. 350 descriptors were obtained at the end of variable / attribute reduction process. Different training and test sets were generated by random data selection methods in VLife MDS 4.4 and the Unicolumn statistics are shown in Figure 1. This result showed that test set was interpolative and derived within the minimum–maximum range of training set. Also, the mean and standard deviation of pIC_{50} values of training and test give insights into the point density distribution and relative difference of mean between the two sets.

MLR, PCR and PLS models are given below as Equation 3, 4 and **5** respectively.

pIC_{50} = 10.1431(±0.9110) Psi1 + 0.0103(±0.0023) +vePotentialSurfaceArea -0.1414(±0.0651) SdsNEindex + 1.0651(±0.2465) SdssCE-index + 0.4231(±0.1251) SdssCcount + 0.0021 ……..……(3)

n = 35, r^2 = 0.7381, r^2_se = 0.4, q^2 = 0.6449, pred_r^2 = 0.5107, z-score_r^2 = 4.61

where,

Psi1 = A measure of hydrogen–bonding propensity of the molecules and/or polar surface area
+vePotentialSurfaceArea = Total van der Waals surface area with positive electrostatic potential of the molecule
SdsNEindex = Electrotopological state indices for number of nitrogen atom connected with two double and one single bond
SdssCE-index = Electrotopological state indices for number of carbon atom connected with one double and two single bonds
SdssCcount = Total number of carbon connected with one double and two single bonds

pIC_{50} = 6.8078 – 0.9604 (SdsNcount) + 0.0505 (Hydrogens Count) ……(4)

n = 32, r^2 = 0.5158, r^2_se = 0.53, q^2 = 0.4665, pred_r^2 = 0.2869, z-score_r^2 = 2.58

where,

SdsNcount = total number of nitrogen connected with one single and one double bond.
HydrogensCount = number of hydrogen atoms in a compound.

pIC_{50} = 4.7593 -1.6082 (SdsNcount) + 0.0056 (SAHydrophobicArea) - 0.1391 (XKMostHydrophobic HydrophilicDistance) + 0.2614 (SsCH3count) …….(5)

n = 32, r^2 = 0.7621, q^2 = 0.49, pred_r^2 = 0.3672, z-score_r^2 = 1.74

where,

SdsNcount = total number of nitrogen connected with one single and one double bond.
SAHydrophobicArea = hydrophobic surface area
XKMostHydrophobicHydrophilicDistance = distance between most hydrophobic and hydrophilic point on the vdW surface
SsCH3count = total number of –CH3 group connected with single bond

Output Window

```
Uni-Column Statistics :

Column Name          Average      Max          Min          StdDev       Sum
pIC50                7.9309       9.3980       6.1800       0.7853       277.5810

=========================================================

=========================================================

Uni-Column Statistics :

Column Name          Average      Max          Min          StdDev       Sum
pIC50                8.4414       8.6990       7.9390       0.2910       42.2070
```

Figure 1. Unicolumn statistics of training and test set respectively

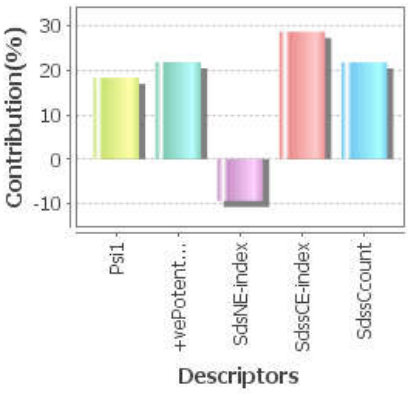

Figure 2. Contribution plot of descriptors

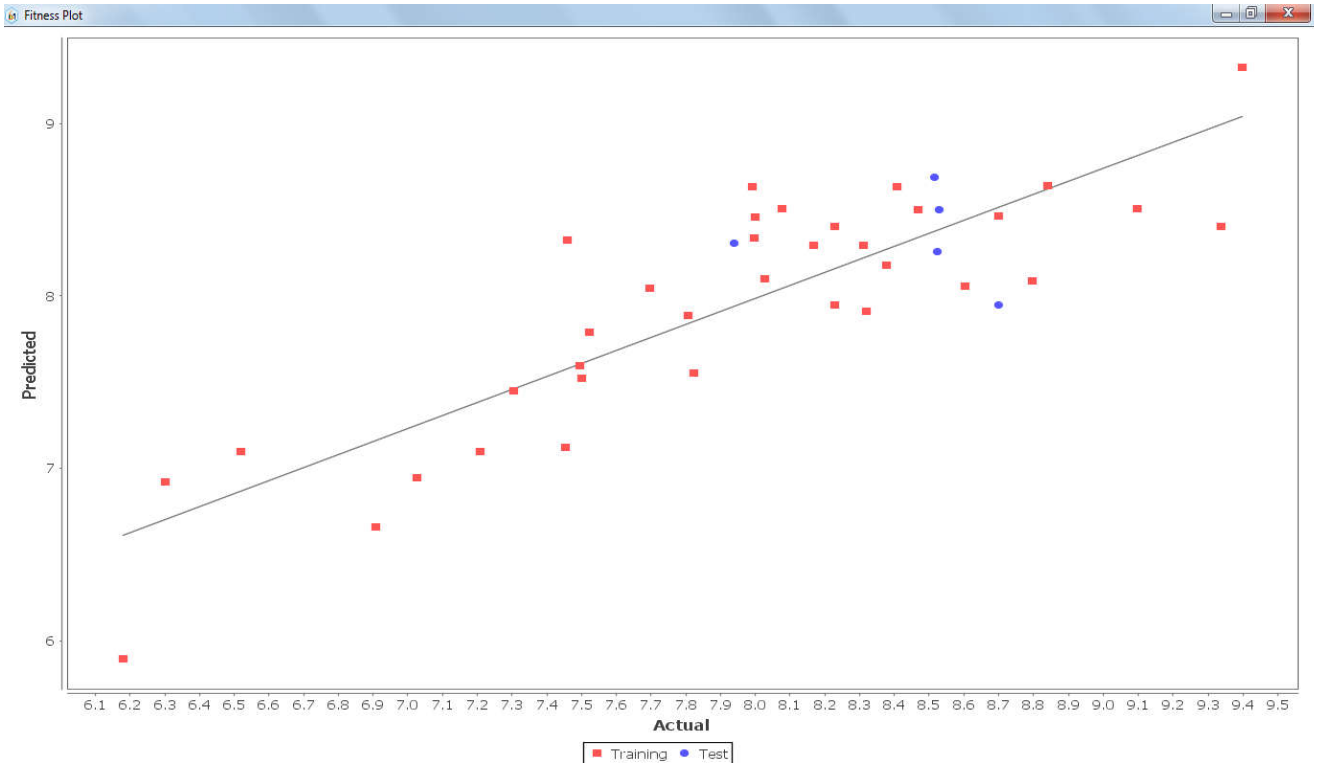

Figure 3. Fitness plot of observed pIC_{50} versus predicted activity of training set and test set compounds

Table 1. The observed and predicted pIC_{50} values for training and test set compounds of best model along with residual values

S. No.	Compound ID	Actual pIC_{50} (μM)	Predicted pIC_{50} (μM) values of MLR based model	Residual (μM)
1	25066547	8.027	8.153	-0.126
2	25227508	8.469	8.472	-0.003
3	25227505	7.991	8.472	-0.481
4	25066548	8.699	8.472	0.227
5	25066355	7.807	8.062	-0.255
6	25227504	9.097	8.472	0.625
7	25066356	7.996	8.107	-0.111
8	25227513	8.409	8.472	-0.063
9	25066549	8.076	8.472	-0.396
10	24779722	7.697	7.971	-0.274
11	6445533	8.229	8.199	0.030
12	444732	9.398	8.062	1.336
13	5311	8.796	7.606	1.190
14	44591988	6.519	6.560	-0.041
15	44138033	7.027	6.651	0.376
16	44138032	6.907	6.515	0.392
17	44138031	6.18	6.515	-0.335
18	9804992	8.377	8.244	0.133
19	44543715	7.523	7.834	-0.311
20	44543714	7.495	7.606	-0.111
21	6918638	7.824	7.606	0.218
22	4996	8.319	7.561	0.758
23	5186	8	8.153	-0.153
24	279980	6.301	7.470	-1.169
25	24951311	7.208	7.652	-0.444
26	24951310	7.305	7.652	-0.347
27	24948910	8.167	7.925	0.242
28	24948909	7.455	7.652	-0.197
29	24948908	8.229	7.743	0.486
30	24948905	7.5	7.789	-0.289
31	50899047	8.31	8.426	-0.116
32	56950025	7.46	8.517	-1.057
33	50899051	8.604	8.199	0.405
34	50898756	9.338	8.517	0.821
35	56950154	8.842	8.654	0.188
36*	50898757	8.524	8.426	0.098
37*	49850262	8.699	8.335	0.364
38*	50898950	7.939	8.381	-0.442
39*	50898670	8.516	8.745	-0.229
40*	50898669	8.529	8.745	-0.216

* represents test set

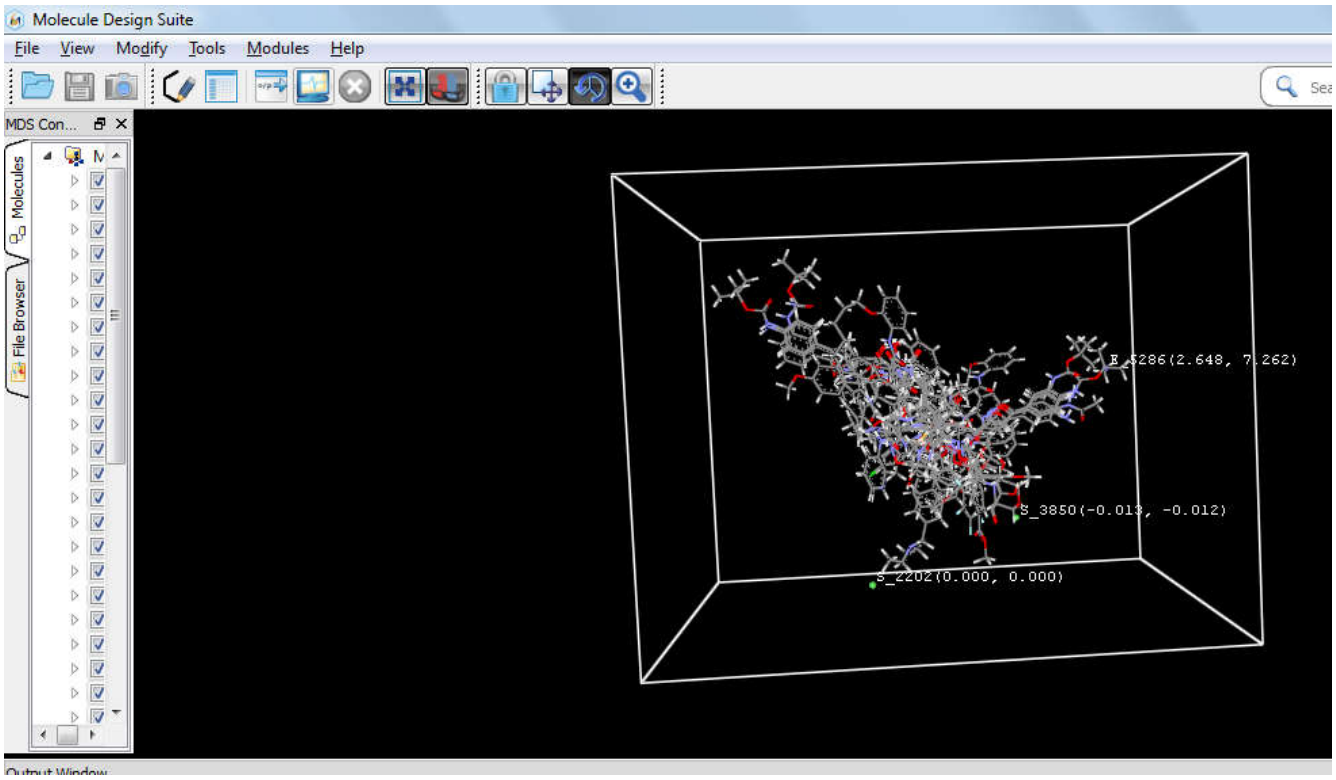

Figure 4. Contour plot of 3D-QSAR model with important steric and electrostatic data points contributing to the model

Comparison of the obtained results indicated the superiority of multiple linear regression model coupled with stepwise forward-backward variable method over other models. MLR model with squared correlation coefficient or coefficient of determination (r^2) = 0.7381 indicates that the model is capable of explaining 73.8% variance in the observed pIC_{50} values. The model has a low standard error of r^2_se = 0.4352. The Cross validated squared correlation coefficient of this model was q^2 = 0.6449 which is greater than 0.5, thus indicating that the model has good internal predictivity in addition to good external predictive power of (pred_r^2 = 0.5107). Z-score, the randomization test result of 4.61 was obtained, showing confidence of ~99.9% that the generated model is not random. Since all the values fall within the acceptable range, MLR model is chosen as the significant QSAR model of all the generated ones. The Contribution chart for the significant model is presented in Figure 2, which gives percentage contribution of descriptors used in deriving the QSAR models. It was found that increase in the value of Psi1, vePotentialSurfaceArea, SdssCE-index and SdssCcount can contribute to the increase in the inhibitory activity of these compounds. Furthermore, the fitness plot between actual and predicted pIC_{50} values was plotted (Figure 3). The observed and predicted pIC_{50} along with residual values are shown in Table 1. Also, kNN-MFA 3D QSAR model was developed based on steric and electrostatic fields using stepwise variable selection method A highly bioactive molecule in the series (molecule 12) was chosen as a reference molecule on which all the other molecules in the data set were aligned, employing template as a basis for the alignment. The data set was split into training (32 compounds) and test set (8 compounds) by random selection method in Vlife MDS 4.4. The quality of the models were assessed by means of internal cross-validation and external validation procedures. The kNN-MFA model showed q^2 of 0.8392, pred_r^2 of 0.4 and k-nearest neighbour of 2. The contour plot of 3D QSAR model is shown in Figure 4.

In the 3D QSAR model, contributing descriptors were E_5286 (2.648, 7.262), S_3850 (-0.013, -0.012). Electrostatic field descriptor E_5286 with positive coefficients represent regions where electron-donating groups are favorable for increase in activity. Steric descriptor S_3850 with negative coefficients indicate that negative steric potential is favourable for increased activity, and hence less bulky substituent group is preferred in that region (Noolvi and Patel, 2013).

Conclusion

The objective of this study was to develop 2D and 3D QSAR models for predicting the inhibitory activity of potent HDAC6 inhibitors. It was evident from our findings that, 2D QSAR model-1 developed by multiple linear regression analysis was found to be statistically significant when compared to all the other models in terms of good internal and external predictive abilities. According to this model, the anti-cancer activity of compounds was influenced by descriptors such as Psi1, +vePotentialSurfaceArea, SdssCE-index, SdssCcount and SdsNE-index. The 2D-QSAR model reported herein provides some interesting insight into understanding the descriptors that contribute significantly for the activity. The grid of 3D model with positive and negative values shows the pattern of substitution for improving the bioactivity of existing compounds. Hence the model proposed in this work can be employed to design new derivatives of HDAC inhibitors or modify the existing ones for improved inhibitory activity.

Acknowledgement

The authors wish to express gratitude to V-life Science Technologies Pvt. Ltd. for providing the license for the software and Dr. Roshan Makam, Head of Department of Biotechnology, PES Institute of Technology, Bangalore for his support throughout the work.

REFERENCES

Ajmani, S., Jadhav, K. and Kulkarni, S.A. 2006. Three-dimensional QSAR using the k-nearest neighbor method and its interpretation. *J. Chem. Inf. Model.,* 46: 24–31.

Aldana Masangkay, G.I. and Sakamoto, K.M. 2011. The role of HDAC6 in cancer. *J. Biomed. Biotechnol*., 2011:875824.

Cherkasov, A., Muratov, E.N., Fourches, D. *et al*. 2014. QSAR Modeling: Where Have You Been? Where Are You Going To?, *J. Med. Chem*., 57:4977–5010.

Dokmanovic, M., Clarke, C. and Marks, P.A. 2013. Histone Deacetylase inhibitors: Overview and Perspectives. *Mol Cancer Res*., 5: 981-989.

Noolvi, M.N. and Patel, H.M. 2013. A comparative QSAR analysis and molecular docking studies of quinazoline derivatives as tyrosine kinase (EGFR) inhibitors: A rational approach to anticancer drug design, *Journal of Saudi Chemical Society.*, 17: 361–379.

Sarankar, S.K. and Pathak, A.K. 2012. QSAR and k-Nearest Neighbor Molecular Field (k-NN MFA) Template Based Analysis of Novel 2-(4-methylsulfonylphenyl) pyrimidine Derivatives as Specific COX-2 Inhibitors. *Journal of Pharmacy Research.,* 5: 1276-1278.

Sharma, M.C. 2015. A Structure–Activity Relationship Study of Naphthoquinone Derivatives as Antitubercular Agents Using Molecular Modeling Techniques. *Interdiscip Sci.,* 7: 346-356.

Sharma, M.C. and Sharma, S. 2015. Molecular modelling study of uracil-based hydroxamic acids-containing histone deacetylase inhibitors, *Arabian journal of chemistry*.

Ventura, C., Latino, D.A. and Martins, F. 2013. Comparison of multiple linear regressions and neural networks based QSAR models for the design of new antitubercular compounds. *Eur. J. Med. Chem.,* 70: 831-845.

Vijayasarathy, S. and Chatterjee, J. 2015. Comparison of MLR, Isotonic Regression and KNN Based QSAR Models for the Prediction of Inhibitory Activity of HDAC6 Inhibitors. *Int. J. Life Sci. Biotech. Pharm. Res*., 4:127-131.

VLife MDS 3.5, Molecular Design Suite, Vlife Sciences Technologies Pvt. Ltd., Pune, India, 2008; software available at http://www.vlifesciences.com/.

Wang, Y., Xiao, J., Suzek, T.O., Zhang, J *et al*. 2012. PubChem's BioAssay Database. *Nucleic Acids Res.*, 40: D400-12.

Wong, K.Y., Mercader, A.G., Saavedra, L.M. *et al*. 2014. SAR analysis on tacrine-related acetylcholinesterase inhibitors. *J Biomed Sci.* 21: 84.

****