

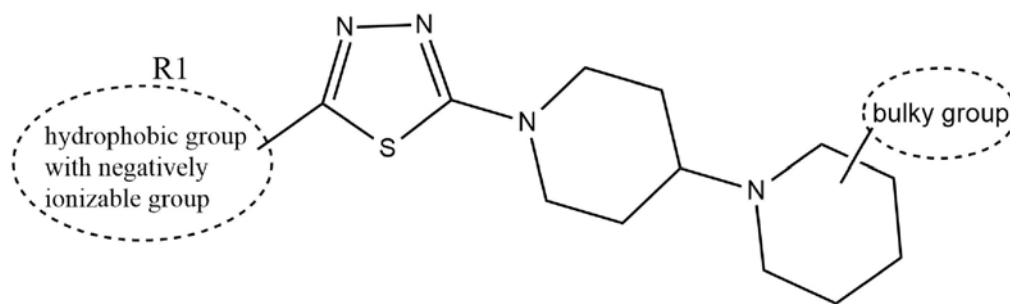
An approach to design potent anti diabetic agents by 3D QSAR studies on substituted thiadiazole derivatives using kNN –MFA methodology

Yogesh Vaishnav^{a*}, Pranita Kashyap^b and Chanchal Deep Kaur^b

^aFaculty of Pharmaceutical Sciences, Shrishankaracharya group of institutions, Junwani, Bhilai

^bShri Rawatpura Sarkar Institute of Pharmacy, Kumhari, Durg ,Chhattisgarh, India

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 2-Piperidinopiperidine thiadiazole derivatives as histamine H3 receptor inhibitors using a molecular design suite (V Life MDS). The present study was performed with 21 compounds (data set) using manual selection and random selection method for the disunion of the data set into training and test set. KNN-MFA approach with stepwise (SW) variable selection forward-backward, Simulated Annealing (SA) and Genetic Algorithms (GA) approaches was used for building the QSAR models. Three predictive models were spawned with Simulated annealing (SA) and Stepwise (SW)-kNN MFA. The most predictive model was spawned by SA-kNN MFA. This model explains well internal ($q^2 = 0.79$) as well as very good external ($\text{Predr}^2 = 0.90$) predictive supremacy of the model. The hydrophobic, electrostatic and steric descriptors at the grid points, H_290, E_348, and S_1351 play an important role in the divulging activity. This model indicates that one hydrophobic, one electrostatic and one steric descriptor is involved. The kNN-MFA contour plots provided a further understanding of the relationship between structural features of substituted Thiadiazole derivatives and their activities, which should be applicable to design newer potential H3 receptor inhibitors.



Proposed Leads

KEYWORDS 3D-QSAR, kNN-MFA, H3 receptor inhibitors, Thiadiazole derivatives.

INTRODUCTION

Diabetes mellitus (DM) is a lingering metabolic disorder, resulting from an insulin deficit. Diabetes is characterized by hyperglycemia, changed the metabolism of carbohydrates, protein, lipids and an enlarged threat of vascular impediments¹.

Compounds comprising 1, 3, 4Thiadiazole nucleuses are of foremost interest due to their inimitable chemical structure and a great range of biological activities². 1, 3, 4 Thiadiazole derivatives have often been evaluated as lead compounds against diabetes³.

*Corresponding author: Email: yogesh446688@gmail.com

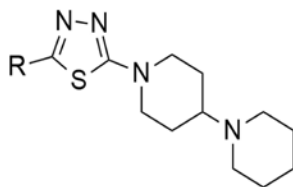
Quantitative structure-activity relationship (QSAR) is a vital part of drug design and discovery, which has become an accepted contrivance for representing a quantitative link between biological activity and descriptors representing physicochemical parameters of the compounds in a series by means of statistical methods⁴.

In recent years, a significant expansion has been made by computational chemistry led new challenges to drug discovery by a rational process. Quantitative structure-activity relationship (QSAR) is a tool for establishing a quantitative relationship between structures and biological activity in a homologous series using statistical

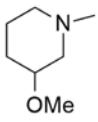
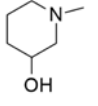
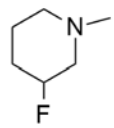
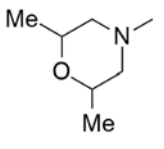
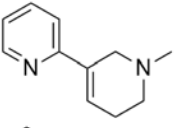
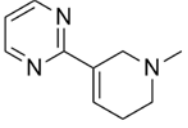
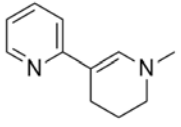
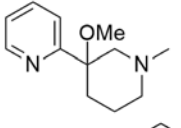
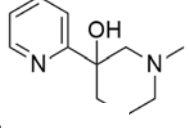
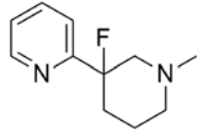
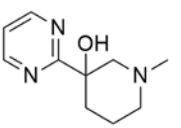
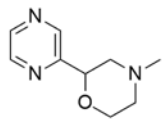
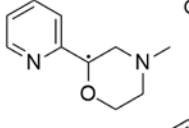
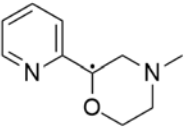
methods, and it helps to envisage the biological behavior of newly intended analogs, which contribute to the drug discovery processes⁵.

The objective of the present study is to seek out for novel 1, 3, 4-Thiadiazole that would show a promise to become useful as Histamine H3 receptor antagonist. A series of 2 piperidinopiperidine thiadiazole derivatives⁶ which was reported as Histamine H3 receptor inhibitors for the treatment of diabetes chosen for QSAR study in order to establish quantitative relationship between physiochemical properties and biological behavior of the compounds using molecular design suite software (VlifeMDS)⁷.

Table 1: Binding affinities of 2-piperidinopiperidine thiadiazole derivatives.



S.No	R	Compound	IC ₅₀ (nM)	pIC ₅₀ (nM)
1		I	49	7.31
2		5a	34	7.46
3		5b	30	7.52
4		5c	120	6.92
5		5d	61	7.21
6		5e	15	7.82
7		5f	75	7.12

8		5g	37	7.43
9		5h	46	7.33
10		5i	66	7.18
11		5j	53	7.27
12		5k	12	7.92
13		5l	14	7.85
14		5m	3.0	8.52
15		5n	11	7.95
16		5o	53	7.27
17		5p	4.0	8.39
18		5q	67	7.17
19		5s	8.0	8.09
20		5t	34	7.46
21		5u	3.0	8.52

MATERIALS AND METHODS

Dataset: In the present study, a data set of 2-piperidinopiperidine thiadiazole derivatives as histamine, H3 receptor inhibitor (21 molecules) 6 has been taken from the literature and used for QSAR studies (Table 1). The reported IC₅₀ values (nM) have been converted to the logarithm scale pIC₅₀ (moles), for QSAR study. The different substituents of all compounds with their actual biological activities are shown in table 1.

METHODOLOGY

Structures of the molecules were drawn in the 2DDrawapp option in Tool menu of QSAR Plus. After that 2D, structures were exported to QSARPlus window (2D structure converted to 3D structure). After the conversion, structures are saved as .mol2 file in the QSARPlus 3D window. Minimization of energy was done by using Merck molecular force field (MMFF) method, which results in the optimization of the geometry of the molecule using the following criteria.

Force field – MMFF (Merck Molecular Force Field)

Charge – MMFF

Max. no. of cycles – 10,000

Convergence criteria (RMS gradients) – 0.01

Dielectric properties

*Distance dependent function

*Constant – 1.0

*Gradient type analysis – 1.0

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

Models were developed by means of k-nearest neighbor molecular field analysis (KNN-MFA) in combination with stepwise (SW) forward backward which comprises (Cross-correlation – 0.5; Term selection– q2; Variance cutoff – 0.0; No. of Max. Neighbors – 5; No. of Min. Neighbors – 2; Select prediction method – Distance-based weighted average), Simulated Annealing(SA) comprises of (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-Auto scaling; No. of Max. Neighbors – 5; No. of Min. Neighbors – 2; Select prediction method – Distance-based weighted average), and Genetic Algorithm consists of (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-Auto scaling; No. of Max. Neighbors – 5; No. of Min. Neighbors – 2; Select prediction method – Distance-based weighted average). SW, SA and GA all three were used for variable selection.

VLife Molecular Design Suite7 (VLifeMDS), allows the user to choose probe, grid size, and grid interval for the generation of descriptors. The variable selection methods along with the corresponding parameters are allowed to be selected, and best possible models are generated by maximizing q2. kNN- MFA requires the appropriate arrangement of a given set of molecules. This is followed by generation of a common rectangular grid around the molecules. Interaction energies like hydrophobic, electrostatic and steric are computed on the lattice points of the grid by means of a methyl probe of charge +1. Energies' interaction values are considered for relationship generation and used as descriptors to decide propinquity between the molecules. The field values at the lattice points were indicated by the word descriptor. The optimal training and test sets were generated by means of random selection and manual selection method. This algorithm allows the construction of training sets, which covers the descriptor space occupied by representative points. After generation of training and test sets, a kNN methodology was applied to the descriptors which were generated over the grid.

MOLECULAR ALIGNMENT:

The selected data set was lined up by using template-based alignment method using the most active molecule 5u as a reference molecule. The alignment of all the molecules on the template is shown in **Fig. 1**. In the template-based alignment method, template structure was defined and used as a basis for alignment of a set of molecules.

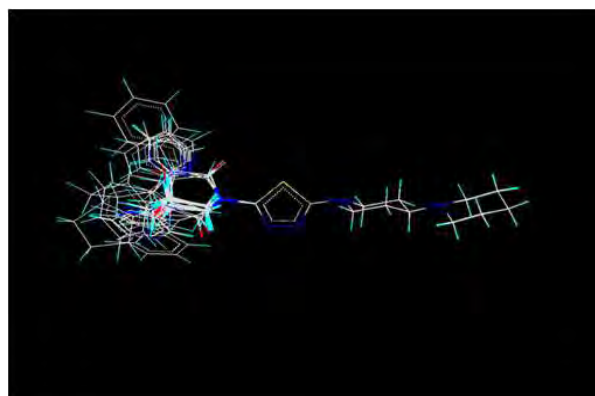


Figure 1: 3D Alignment of molecules

MODEL DEVELOPMENT:

The activity data were subjected to kNN-MFA method for model building.

Data Selection: Data's selection was done by random and manual selection method in which the whole data was divided into training and test sets through which external validation of qsar models was evaluated.

The training set is used to develop the QSAR model for which biological activity data are known. The test set is used to challenge the QSAR model developed based on the training set to assess the predictive efficiency of the model which is not included in model generation. Descriptors were chosen as independent variable and biological activity is chosen as a dependent variable.

For the creation of training and test set following methods were used.

- Manual selection method
- Random selection method

Random selection: In order to construct and validate the QSAR models, both internally and externally, the data sets were divided into a training set of the total data set and test sets of total data set in a random manner. The

percentage of training set and test set should be 85%, 80%, 75%, 70% and 15%, 20%, 25% and 30% respectively. In each selection, 10 trials were run.

Manual data selection: Whole range of biological activities was sorted in ascending and descending manner and every 3rd, 4th, 5th, 6th, 7th, 8th and 9th molecule was assigned to the test set.

RESULTS AND DISCUSSION

A series of 2-piperidinopiperidine thiadiazole derivatives was selected and subjected for generation of statistically significant models. Different QSAR models were developed using random selection and manual selection of data set. Training and test set were selected if they follow unicolon statistics shown in **Table 2**.

Model 1 was obtained by manual data selection for test set and simulated annealing as a variable selection method. It has been very good internal and the external predictive ability of ~79% and ~91% correspondingly. According to this model, the field grid points H₂₉₀, E₃₄₈, and S₁₃₅₁ representing hydrophobic, electrostatic and steric fields respectively play a significant role in the determination of biological activity. H₂₉₀ with its positive

Table 2: Unicolon statistic of best models

Model no	Data set	Average	Max	Min	Std dev	Sum
1	Training	7.6	8.5	6.9	0.52	114.6
	Test	7.5	8.0	7.1	0.36	45.2
2	Training	7.6	8.5	6.9	0.53	115.2
	Test	7.4	7.8	7.2	0.21	44.5
3	Training	7.5	8.5	6.9	0.48	113.2
	Test	7.6	8.4	7.3	0.45	46.5

Model 1 : Results of 3D-QSAR Analysis using k NN method (k nearest neighbor) by Manual selection method was shown in Table 4. For generation of model 1 following parameters were used "Manual Selection-Ascending-KNN-Simulated annealing-(Trial 1)-Every 3rd Molecule".

Table 3: Results of 3D-QSAR Analysis using k NN method (k nearest neighbor) by Manual selection method.

Model 1:- Manual Selection-Ascending-KNN-Simulated annealing-(Trial 1)-Every 3rd Molecule

S.No	Trial	Test set	KNN Method result
1	1	5h	k Nearest Neighbour= 2
		5j	n = 15
		5l	Degree of freedom = 11
		5q	q ₂ = 0.79
		5s	q _{2_se} = 0.23
		5t	Predr ₂ = 0.90
			pred_r _{2se} = 0.11

K Nearest Neighbor = 2, n = 15, Degree of freedom = 11, q₂ = 0.79, q_{2_se} = 0.23, Predr₂ = 0.90, pred_r_{2se} = 0.11



range (0.230523 0.264163) and position away from indicates the need of the hydrophobic group at R1. E_348 is near R1 and has a negative range (-7.07331 -3.53007) which directs the need of the negatively ionizable group at this position. S_1351 is located near terminal piperidine moiety and has a negative range (-0.014381 -0.014372) thus signifying the

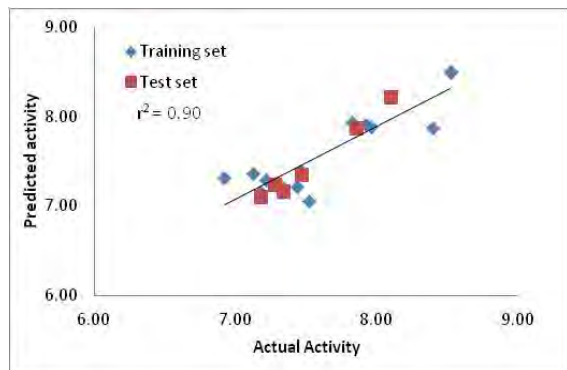


Fig 2: Graph of actual and predicted biological activity for training and test set (Model 1).

need of the bulky group for favorable biological activity. The Graph of actual and predicted biological activity for training and test set of Model 1 and 3D-alignment of molecules with the important steric and electrostatic points contributing Model 1 with ranges of values shown in figure 2&3.

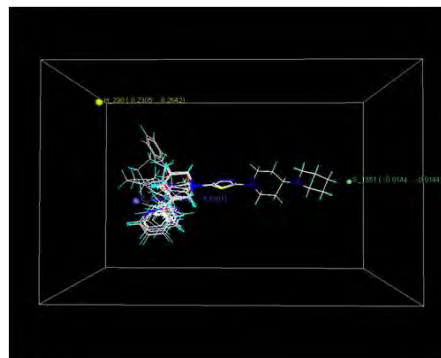


Fig 3: MFA result (Show points): 3D-alignment of molecules with the important steric and electrostatic points contributing [Model 1] with ranges of values shown in parenthesis.

Model 2 : Results of 3D-QSAR Analysis using k NN method (k nearest neighbor) by Random selection method - stepwise regression method was shown in Table 4. For generation of model 2 following parameters were used “Random selection-75%-KNN-Step Wise – (Trial 5)”.

Table 4: Results of 3D-QSAR Analysis using k NN method (k nearest neighbor) by Random selection method - stepwise regression method

Model 2:- Random selection-75%-KNN-Step Wise – (Trial 5)

S.No	Trial	Test set	KNN Method result
1	5	5a	k Nearest Neighbour= 3
		5d	n = 15
		5e	Degree of freedom = 12
		5g	q2 = 0.80
		5h	q2_se = 0.24
		I	Predr2 = 0.64
			pred_r2se = 0.20

k Nearest Neighbour= 3, n = 15, Degree of freedom = 12, q² = 0.80, q²_{se} = 0.24, Predr² = 0.64, pred_r²_{se} = 0.20

Model 2 was obtained by random selection method for test set and stepwise as a variable selection method. It has internal and external predictive ability of ~80% and ~65% respectively. According to this model, the field grid points E_348 and S_475 representing electrostatic and steric fields respectively play a significant role in the determination of biological activity. E_348 is near R1 and has a negative range (-7.07331 -3.53007) which directs the need of a negatively ionizable group at this position.

Furthermore, S_475 is located near R1 and has a negative range (-0.424139 -0.397217) thus signifying the need of a fewer bulky group for favorable biological activity. The Graph of actual and predicted biological activity for training and test set of Model 2 and 3D-alignment of molecules with the important steric and electrostatic points contributing Model 2 with ranges of values was shown in figure 4&5.

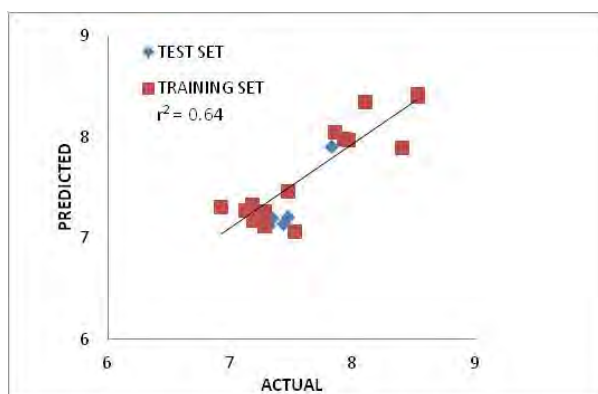


Fig 4: Graph of actual and predicted biological activity for training and test set (Model 2)

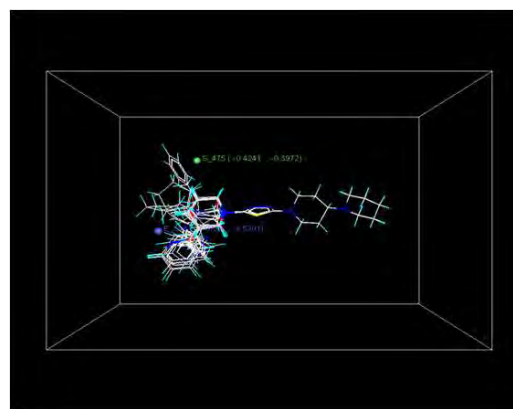


Fig 5: MFA result (Show points): 3D-alignment of molecules with the important steric and electrostatic points contributing [Model 2] with ranges of values shown in parenthesis

Model 3 : Results of 3D-QSAR Analysis using k NN method (k nearest neighbor) by Random selection method –stepwise regression method was shown in Table 5. For generation of model 2 following parameters were used “Random selection-75%-KNN-Step Wise – (Trial 8)”.

Table 5: Results of 3D-QSAR Analysis using k NN method (k nearest neighbor) by Random selection method – stepwise regression method.

Model 3:- Random selection-75%-KNN-Step Wise – (Trial 8)

S.No	Trial	Test set	KNN Method result
1	8	5g	k Nearest Neighbour= 4
		5h	n= 15
		5n	Degree of freedom = 11
		5p	q2 = 0.93
		5s	q2_se = 0.12
		I	Predr2 = 0.56
			pred_r2se = 0.33

Models 3 have obtained by random selection method for test set and stepwise as a variable selection method. It has internal and external predictive ability of ~93% and ~56% respectively. According to this model, the field grid points E_348, S_465 and S_476 representing electrostatic and steric fields respectively play a significant role in determination of biological activity. E_348 is near R1 and has a negative range (-7.07331 -2.91707) which directs the need of a negatively ionizable group at this position. S_465, and S_476 was located near R1 and has positive range (30 30) and (0.666979 1.22596) respectively thus signifying the need of a more bulky group for favorable biological activity. The graph of actual and predicted biological activity for training and test set of Model 3 and 3D-alignment of molecules with the important steric and electrostatic points contributing Model 2 with ranges of values was shown in figure 6&7.

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 2piperidinopiperidinethiadiazole derivatives as histamine H3 receptor inhibitors using a molecular design suite (VLifeMDS). The study was performed with 21 compounds (data set) using manual and random data selection method for the division of data set into training and test set. KNN-MFA methodology with stepwise variable selection forward and backward. Simulated Annealing and Genetic Algorithms methods were used for building the QSAR models. Three predictive models were generated in which one model by manual SA- kNN-MFA, and the other two models were generated by SW-kNN- MFA. The most predictive model was generated by kNN (Simulated annealing) using manual data

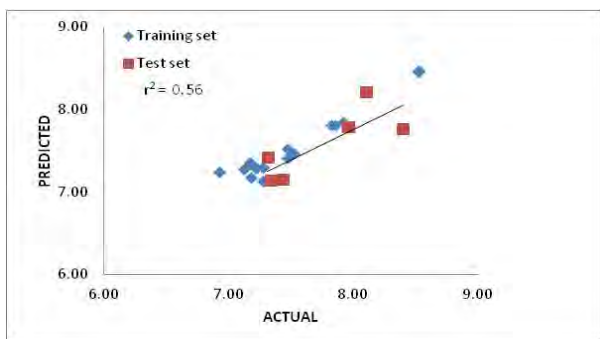


Fig 6: Graph of actual and predicted biological activity for training and test set (Model 3)

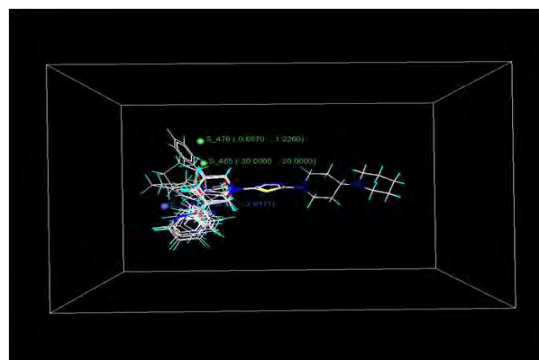


Fig 7: MFA result (Show points): 3D-alignment of molecules with the important steric and electrostatic points contributing [Model 3] with ranges of values shown in parenthesis.

The actual and predicted activity of training and test set of best models were shown in Table 6 to 11.

Table 6: Actual and predicted values of training set of model 1

S.no.	Compound	Actual	Predicted
1.	5a	7.46	7.39
2.	5b	7.52	7.05
3.	5c	6.92	7.31
4.	5d	7.21	7.28
5.	5e	7.82	7.94
6.	5f	7.12	7.36
7.	5g	7.43	7.21
8.	5i	7.18	7.15
9.	5k	7.92	7.90
10.	5m	8.52	8.50
11.	5n	7.95	7.88
12.	5o	7.27	7.25
13.	5p	8.39	7.87
14.	5u	8.52	8.49
15.	I	7.31	7.25

Table 7: Actual and predicted values of test set of model 1

S.no.	Compound	Actual	Predicted
1.	5h	7.337	7.16
2.	5j	7.276	7.24
3.	5l	7.854	7.86
4.	5q	7.174	7.10
5.	5s	8.097	8.22
6.	5t	7.469	7.35

Table 8: Actual and predicted values of training set of model 2

S.no.	Compound	Actual	Predicted
1.	5b	7.523	7.07
2.	5c	6.921	7.31
3.	5f	7.125	7.27
4.	5i	7.18	7.17
5.	5j	7.276	7.12
6.	5k	7.921	7.98
7.	5l	7.854	8.05
8.	5m	8.523	8.42
9.	5n	7.959	7.97
10.	5o	7.276	7.26
11.	5p	8.398	7.90
12.	5q	7.174	7.33
13.	5s	8.097	8.35
14.	5t	7.469	7.47
15.	5u	8.523	8.40

Table 9: Actual and predicted values of test set of model 2

S.no.	Compound	Actual	Predicted
1.	5a	7.469	7.21
2.	5d	7.215	7.30
3.	5e	7.824	7.91
4.	5g	7.432	7.14
5.	5h	7.337	7.20
6.	I	7.31	7.12

Table 10: Actual and predicted values of training set of model 3

S.no.	Compound	Actual	Predicted
1.	5a	7.469	7.52
2.	5b	7.523	7.46
3.	5c	6.921	7.24
4.	5d	7.215	7.28
5.	5e	7.824	7.81
6.	5f	7.125	7.27
7.	5i	7.18	7.17
8.	5j	7.276	7.12
9.	5k	7.921	7.84
10.	5l	7.854	7.81
11.	5m	8.523	8.47
12.	5o	7.276	7.29
13.	5q	7.174	7.34
14.	5t	7.469	7.40
15.	5u	8.523	8.46

Table 11: Actual and predicted values of test set of model 3

S.no.	Compound	Actual	Predicted
1.	5g	7.432	7.14
2.	5h	7.337	7.14
3.	5n	7.959	7.78
4.	5p	8.398	7.76
5.	5s	8.097	8.21
6.	I	7.31	7.41

Table 12: Statistical significant models (best 3) generated.

Parameters	Model 1	Model 2	Model 3
	Manual -ASC – KNN-SA (Trial 1)	RAN-75%-KNN- SW-(Trial 5)	RAN-75%-KNN- SW-(Trial 8)
Training set (n)	15	15	15
Test set size	06	06	06
Test set	5h,5j,5l,5q,5s,5t	5a,5d,5e,5g,5h,I	5g,5h,5n,5p,5s,I
Degree of freedom	11	12	11
q ²	0.79	0.80	0.93
q ² _se	0.23	0.24	0.12
Predr ²	0.90	0.64	0.56
pred_r ² se	0.11	0.20	0.33
Descriptor	H_290E_348S_1351	E_348S_475	E_348S_465S_476

selection method. This model explains good internal ($q^2 = 0.79$) as well as very good external ($\text{predr}^2 = 0.90$) predictive supremacy of the model. The hydrophobic, electronic and steric descriptors at the grid points, H_290, E_348, and S_1351 play an important role in the divulging activity. This model indicates that one hydrophobic, one electrostatic and one steric descriptor was involved. The kNN-MFA contour plots provided a further understanding of the relationship between structural features of substituted thiadiazole derivatives and their activities, which should be applicable to design newer potential H3 receptor inhibitors. The kNN-MFA contour plots provided a further understanding of the relationship between structural features of substituted thiadiazole derivatives and their activities, which should be applicable to design newer potential H3 receptor inhibitors.

ACKNOWLEDGEMENT

I would like to express my appreciation to Dr. Sanmati Kumar Jain and Mr. Piyush Ghode of SLT Institute of

Pharmaceutical Sciences, Guru Ghasidas Vishwavidyalaya Koni Bilaspur for their help and support during my work. Without their valuable assistance, this work would not have been completed.

REFERENCES

- [1] FSK .Barar; *Essentials of Pharmacotherapeutics*”, S Chand, New Delhi (2007), 340.
- [2] Lincy Joseph, Mathew George, Prabha Mathews, *Journal of Pharmaceutical, Chemical and Biological Sciences.*, **3** (2015), 329-345.
- [3] Prasanna A Datar* and Tejashree A Deokule, *Medicinal Chemistry.*, **4** (2014), 390-399.
- [4] M.M.C. Ferreira, *J. Braz. Chem. Soc.*, **13** (2002) 742
- [5] Ferreira MMC. Multivariate QSAR. *J BrazChem Soc.* **2002**; 13:742.
- [6] AshwinU.Rao, *ACS Medicinal chemistry letters.*, **3** (2012), 198-202.
- [7] VLifeMDS 3.5, Molecular Design Suite, Vlife SciencesTechnologies Pvt. Ltd., Pune, India (2009), www.vlifesciences.com.

